Camera Self-Calibration: Theory and Experiments

O.D. Faugeras[1], *Q.-T. Luong*[1] and *S.J. Maybank*[2]

[1] INRIA, 2004 Route des Lucioles, 06560 Valbonne, France
[2] GEC Hirst Research Centre, East Lane, Wembley, Middlesex, HA9 7PP UK

Abstract. The problem of finding the internal orientation of a camera (camera calibration) is extremely important for practical applications. In this paper a complete method for calibrating a camera is presented. In contrast with existing methods it does not require a calibration object with a known 3D shape. The new method requires only point matches from image sequences. It is shown, using experiments with noisy data, that it is possible to calibrate a camera just by pointing it at the environment, selecting points of interest and then tracking them in the image as the camera moves. It is not necessary to know the camera motion.

The camera calibration is computed in two steps. In the first step the epipolar transformation is found. Two methods for obtaining the epipoles are discussed, one due to Sturm is based on projective invariants, the other is based on a generalisation of the essential matrix. The second step of the computation uses the so-called Kruppa equations which link the epipolar transformation to the image of the absolute conic. After the camera has made three or more movements the Kruppa equations can be solved for the coefficients of the image of the absolute conic. The solution is found using a continuation method which is briefly described. The intrinsic parameters of the camera are obtained from the equation for the image of the absolute conic.

The results of experiments with synthetic noisy data are reported and possible enhancements to the method are suggested.

1 Introduction

Camera calibration is an important task in computer vision. The purpose of the camera calibration is to establish the projection from the 3D world coordinates to the 2D image coordinates. Once this projection is known, 3D information can be inferred from 2D information, and vice versa. Thus camera calibration is a prerequisite for any application where the relation between 2D images and the 3D world is needed. The camera model considered is the one most widely used. It is the pinhole: the camera is assumed to perform a perfect perspective transformation. Let $[su, sv, s]$ be the image coordinates, where s is a non-zero scale factor. The equation of the projection is

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \mathbf{A} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{G} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \equiv \mathbf{M} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where X, Y, Z are world coordinates, $\mathbf{A}$ is a 3×3 transformation matrix accounting for camera sampling and optical characteristics and $\mathbf{G}$ is a 4×4 displacement matrix

accounting for camera position and orientation. Projective coordinates are used in (1) for the image plane and for 3D space. The matrix **M** is the perspective transformation matrix, which relates 3D world coordinates and 2D image coordinates. The matrix **G** depends on six parameters, called extrinsic: three defining a rotation of the camera and three defining a translation of the camera. The matrix **A** depends on a variable number of parameters, according to the sophistication of the camera model. These are the intrinsic parameters. There are five of them in the model used here. It is the intrinsic parameters that are to be computed.

In the usual method of calibration [4] [12] a special object is put in the field of view of the camera. The 3D shape of the calibration object is known, in other words the 3D coordinates of some reference points on it are known in a coordinate system attached to the object. Usually the calibration object is a flat plate with a regular pattern marked on it. The pattern is chosen such that the image coordinates of the projected reference points (for example, corners) can be measured with great accuracy. Using a great number of points, each one yielding an equation of the form (1), the perspective transformation matrix **M** can be estimated. This method is widely used. It yields a very accurate determination of the camera parameters, provided the calibration pattern is carefully set. The drawback of the method is that in many applications a calibration pattern is not available. Another drawback is that it is not possible to calibrate on-line when the camera is already involved in a visual task. However, even when the camera performs a task, the intrinsic parameters can change intentionally (for example adjustment of the focal length), or not (for example mechanical or thermal variations).

The problem of calibrating the extrinsic parameters on-line has already been addressed [11]. The goal of this paper is to present a calibration method that can be carried out using the same images required for performing the visual task. The method applies when the camera undergoes a series of displacements in a rigid scene. The only requirement is that the machine vision is capable of establishing correspondences between points in different images, in other words it can identify pairs of points, one from each image, that are projections of the same point in the scene.

Many methods for obtaining matching pairs of points in two images are described in the literature. For example, points of interest can be obtained by corner and vertex detection [1] [5]. Matching is then done by correlation techniques or by a tracking method such as the one described in [2].

2 Kruppa's Equations and Self-Calibration

A brief introduction is given to the theory underlying the calibration method. A longer and more detailed account is given in [9].

2.1 Derivation of Kruppa's Equations

Kruppa's equations link the epipolar transformation to the image w of the absolute conic Ω. The conic w determines the camera calibration, thus the equations provide a way of deducing the camera calibration from the epipolar transformations associated with a sequence of camera motions. Three epipolar transformations, arising from three different camera motions, are enough to determine w and hence the camera calibration uniquely.

The absolute conic is a particular conic in the plane at infinity. It is closely associated with the Euclidean properties of space. The conic Ω is invariant under rigid motions and

under uniform changes of scale. In a Cartesian coordinate system $[x_1, x_2, x_3, x_4]$ for the projective space $\mathbb{P}^3$ the equations of Ω are

$$x_4 = 0 \qquad\qquad x_1^2 + x_2^2 + x_3^2 = 0$$

The invariance of Ω under rigid motions ensures that w is independent of the position and orientation of the camera. The conic $w = \mathbf{M}(\Omega)$ thus depends only on the matrix $\mathbf{A}$ of intrinsic parameters. The converse is also true [9] in that w determines the intrinsic parameters.

Let the camera undergo a finite displacement and let k be the line joining the optical centre of the camera prior to the motion to the optical centre of the camera after the motion. Let $\mathbf{p}$ and $\mathbf{p}'$ be the epipoles associated with the displacement. The epipole $\mathbf{p}$ is the projection of k into the first image and $\mathbf{p}'$ is the projection of k into the second image. Let Π be a plane containing k. Then Π projects to lines l and l' in the first and second images respectively. The epipolar transformation defines a homography from the lines through $\mathbf{p}$ to the lines through $\mathbf{p}'$ such that l and l' correspond. The symbol $\overline{\wedge}$ is used for the homographic correspondence, $l \overline{\wedge} l'$.

If Π is tangent to Ω then l is tangent to w and l' is tangent to the projection w' of Ω into the second image. The conic w is independent of the camera position thus $w = w'$. It follows that the two tangents to w from $\mathbf{p}$ correspond under the epipolar transformation to the two tangents to w from $\mathbf{p}'$. The condition that the epipolar lines tangent to w correspond gives two constraints linking the epipolar transformation with w. Kruppa's equations are an algebraic version of these constraints.

Projective coordinates $[y_1, y_2, y_3]$ are chosen in the first image. Two triples of coordinates $[y_1, y_2, y_3]$ and $[u_1, u_2, u_3]$ specify the same image point if and only if there exists a non-zero scale factor s such that $y_i = su_i$ for $i = 1, 2, 3$. The epipolar lines are parameterised by taking the intersection of each line with the fixed line $y_3 = 0$. Let $\langle \mathbf{p}, \mathbf{y} \rangle$ be the line through the two points $\mathbf{p}$ and $\mathbf{y}$. A general point $\mathbf{x}$ is on $\langle \mathbf{p}, \mathbf{y} \rangle$ if and only if $(\mathbf{p} \times \mathbf{y}).\mathbf{x} = 0$. Let $\mathbf{D}$ be the matrix of the dual conic to w. It follows from the definition of $\mathbf{D}$ that $\langle \mathbf{p}, \mathbf{y} \rangle$ is tangent to w if and only if it lies on the dual conic,

$$(\mathbf{p} \times \mathbf{y})^\top \mathbf{D} (\mathbf{p} \times \mathbf{y}) = 0 \tag{2}$$

The entries of $\mathbf{D}$ are defined to agree in part with the notation of Kruppa,

$$\mathbf{D} = \begin{bmatrix} -\delta_{23} & \delta_3 & \delta_2 \\ \delta_3 & -\delta_{13} & \delta_1 \\ \delta_2 & \delta_1 & -\delta_{12} \end{bmatrix} \tag{3}$$

There are six parameters δ_i, δ_{ij} in (3), but $\mathbf{D}$ is determined by w only up to a scale factor. After taking the scale factor into account $\mathbf{D}$ has five degrees of freedom. On setting $y_3 = 0$ and on using (3) to substitute for $\mathbf{D}$ in (2) it follows that

$$A_{11} y_1^2 + 2A_{12} y_1 y_2 + A_{22} y_2^2 = 0 \tag{4}$$

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

$$\begin{aligned} A_{11} &= -\delta_{13} p_3^2 - \delta_{12} p_2^2 - 2\delta_1 p_2 p_3 \\ A_{12} &= \delta_{12} p_1 p_2 - \delta_3 p_3^2 + \delta_2 p_2 p_3 + \delta_1 p_1 p_3 \\ A_{22} &= -\delta_{23} p_3^2 - \delta_{12} p_1^2 - 2\delta_2 p_1 p_3 \end{aligned} \tag{5}$$

An equation similar to (4) is obtained from the condition that the epipolar line $\langle \mathbf{p}', \mathbf{y}' \rangle$ in the second image is tangent to w,

$$A'_{11}y'^2_1 + 2A_{12}y'_1y'_2 + A'_{22}y'^2_2 = 0 \tag{6}$$

The coefficients A'_{11}, A'_{12}, A'_{22} are obtained from (5) on replacing the coordinates p_i of $\mathbf{p}$ with the coordinates p_i of $\mathbf{p}'$. The coordinate y'_3 is set equal to zero.

The epipolar transformation induces a bilinear transformation N from the line $y_3 = 0$ to the line $y'_3 = 0$. If $\mathbf{y} = [y_1, y_2, 0]^T$ and $\mathbf{y}' = [y'_1, y'_2, 0]^T$ then $\langle \mathbf{p}, \mathbf{y} \rangle \overline{\wedge} \langle \mathbf{p}', \mathbf{y}' \rangle$ if and only if $\mathbf{y}' = N\mathbf{y}$. Let $\tau = y_2/y_1$, $\tau' = y'_2/y'_1$. Then the transformation N is equivalent to

$$\tau' = \frac{a\tau + b}{c\tau + d} \tag{7}$$

The parameters a, b, c, d can be easily computed up to a scale factor from the two epipoles $\mathbf{p}$,$\mathbf{p}'$ and a set of point matches $\mathbf{q}_i \leftrightarrow \mathbf{q}'_i$, $1 \leq i \leq n$. A linear least square procedure based on (7) is used. The ith image correspondence gives an equation (7) with

$$\tau_i = \frac{p_3q_{i2} - p_2q_{i3}}{p_3q_{i1} - p_1q_{i3}} \qquad \tau'_i = \frac{p'_3q'_{i2} - p'_2q'_{i3}}{p'_3q'_{i1} - p'_1q'_{i3}} \tag{8}$$

Once a, b, c, d have been found (4), (6) and (7) yield

$$\begin{aligned} A_{11} + 2A_{12}\tau + A_{22}\tau^2 &= 0 \\ A'_{11}(b\tau + c)^2 + 2A'_{12}(b\tau + c)(\tau + a) + A'_{22}(\tau + a)^2 &= 0 \end{aligned} \tag{9}$$

Each equation (9) is quadratic in τ. The two equations have the same roots, namely the values of τ for which $\langle \mathbf{p}, \mathbf{y} \rangle$ is tangent to w. It follows that one equation is a scalar multiple of the other. Kruppa's equations are obtained by equating ratios of coefficients,

$$\begin{aligned} A_{12}(A'_{22}a^2 + A'_{11}c^2 + 2A'_{12}ac) - (A'_{12}c + A'_{22}a + A'_{11}bc + A'_{12}ab)A_{11} &= 0 \\ A_{22}(A'_{22}a^2 + A'_{11}c^2 + 2A'_{12}ac) - (2A'_{12}b + A'_{22} + A'_{11}b^2)A_{11} &= 0 \end{aligned}$$

2.2 Kruppa's Equations for Two Camera Motions

Two camera motions yield two epipolar transformations and hence four constraints on the image w of the absolute conic. The conic w depends on five parameters, thus the conics compatible with the four constraints form a one dimensional family c. The family c is an algebraic curve which parameterises the camera calibrations compatible with the two epipolar transformations.

An algebraic curve can be mapped from one projective space to another using transformations defined by polynomials. A linear transformation is a special case in which the defining polynomials have degree one. One approach to the theory of algebraic curves is to regard each transformed curve as a different representation of the same underlying algebraic object. For example a conic, a plane cubic with a node and a cubic space curve can all be obtained by applying polynomial transformations to the projective line $\mathbb{P}^1$. Each curve is a different representation of $\mathbb{P}^1$, even though the three curves appear to be very different.

The properties of c are obtained in [9]. It is shown that c can be represented as an algebraic curve of degree seven in $\mathbb{P}^3$ or alternatively as an algebraic curve of degree six in $\mathbb{P}^2$. The representation of c as a curve in $\mathbb{P}^2$ is obtained as follows.

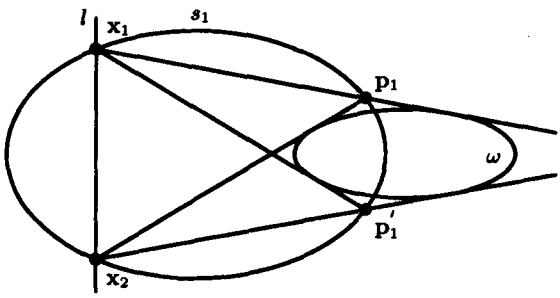

Fig. 1. Construction of the dual curve g

Let $\mathbf{p}_1$, $\mathbf{p}_1'$ be the two epipoles for the first motion of the camera. The epipolar transformation is defined by the Steiner conic s_1 through $\mathbf{p}_1$ and $\mathbf{p}_1'$; two epipolar lines $\langle \mathbf{p}_1, \mathbf{y} \rangle$ and $\langle \mathbf{p}_2, \mathbf{y} \rangle$ correspond if and only if $\mathbf{y}$ is a point of s_1. The two tangents from $\mathbf{p}_1$ to w cut s_1 at points $\mathbf{x}_1$, $\mathbf{x}_2$ as illustrated in Fig. 1. The chord $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ of s_1 corresponds to the point $\mathbf{x}_1 \times \mathbf{x}_2$ in the dual of the image plane. The point $\mathbf{x}_1 \times \mathbf{x}_2$ lies on a curve g which is an algebraic transformation of c. It is shown in [9] that g is of degree six and genus four. The point $\mathbf{p}_1 \times \mathbf{p}_1'$ of g corresponding to the line $\langle \mathbf{p}_1, \mathbf{p}_1' \rangle$ in the image plane is a singular point of multiplicity three. The curve g has three additional singular points, each of order two. An algorithm for obtaining these three singular points is described in [9]. The algorithm produces a cubic polynomial equation in one variable, the roots of which yield the three singular points.

Three camera displacements yield six conditions on the camera calibration. This is enough to determine the camera calibration uniquely.

3 Computing the Epipoles

Two different methods for computing the epipoles are described.

3.1 Sturm's Method

The epipoles and the epipolar transformations can be computed by a method due to Hesse [6] and nicely summarized by Sturm in [10]. Sturm's method yields the epipoles compatible with seven image correspondences.

Let $\mathbf{q}_i \leftrightarrow \mathbf{q}_i'$, $1 \leq i \leq n$, be a set of image correspondences. Then $\mathbf{p}$, $\mathbf{p}'$ are possible epipoles if and only if

$$\langle \mathbf{p}, \mathbf{q}_i \rangle \overline{\wedge} \langle \mathbf{p}', \mathbf{q}_i' \rangle \qquad 1 \leq i \leq n \tag{10}$$

The pencil of lines through $\mathbf{p}$ is parameterised by the points of the line $\langle \mathbf{q}_1, \mathbf{q}_2 \rangle$. Let $\langle \mathbf{p}, \mathbf{q} \rangle$ be any line through $\mathbf{p}$. Then $\langle \mathbf{p}, \mathbf{q} \rangle$ meets $\langle \mathbf{q}_1, \mathbf{q}_2 \rangle$ at the point $\mathbf{x}$ defined by

$$\begin{aligned} \mathbf{x} &= (\mathbf{p} \times \mathbf{q}) \times (\mathbf{q}_1 \times \mathbf{q}_2) \\ &= [(\mathbf{p} \times \mathbf{q}).\mathbf{q}_2]\mathbf{q}_1 - [(\mathbf{p} \times \mathbf{q}).\mathbf{q}_1]\mathbf{q}_2 \end{aligned}$$

The line $\langle \mathbf{p}, \mathbf{q}_i \rangle$ is assigned the inhomogeneous coordinate θ_i defined by

$$\theta_i = \frac{(\mathbf{p} \times \mathbf{q}_i).\mathbf{q}_2}{(\mathbf{p} \times \mathbf{q}_i).\mathbf{q}_1} = \frac{(\mathbf{q}_i \times \mathbf{q}_2).\mathbf{p}}{(\mathbf{q}_i \times \mathbf{q}_1).\mathbf{p}}$$

It follows that $\theta_1 = \infty$ and $\theta_2 = 0$. Let θ_q be the inhomogeneous coordinate of a fifth line $\langle \mathbf{p}, \mathbf{q} \rangle$. The cross ratio τ of the four lines through $\mathbf{p}$ with inhomogeneous coordinates θ_1, θ_2, θ_3, θ_q is

$$\begin{aligned}\tau &= \left(\frac{\theta_1 - \theta_3}{\theta_2 - \theta_3}\right) \Big/ \left(\frac{\theta_1 - \theta_q}{\theta_2 - \theta_q}\right) \\ &= \frac{\theta_q}{\theta_3} \\ &= \left(\frac{(\mathbf{p} \times \mathbf{q}).\mathbf{q}_2}{(\mathbf{p} \times \mathbf{q}).\mathbf{q}_1}\right) \left(\frac{(\mathbf{p} \times \mathbf{q}_3).\mathbf{q}_1}{(\mathbf{p} \times \mathbf{q}_3).\mathbf{q}_2}\right) \end{aligned} \tag{11}$$

It follows from (10) that τ is equal to the cross ratio of the lines $\langle \mathbf{p}', \mathbf{q}_i' \rangle$, $1 \leq i \leq 3$ and $\langle \mathbf{p}', \mathbf{q}' \rangle$ in the second image. On equating the two cross ratios the following equation is obtained.

$$\left(\frac{(\mathbf{p} \times \mathbf{q}).\mathbf{q}_2}{(\mathbf{p} \times \mathbf{q}).\mathbf{q}_1}\right) \left(\frac{(\mathbf{p} \times \mathbf{q}_3).\mathbf{q}_1}{(\mathbf{p} \times \mathbf{q}_3).\mathbf{q}_2}\right) = \left(\frac{(\mathbf{p}' \times \mathbf{q}').\mathbf{q}_2'}{(\mathbf{p}' \times \mathbf{q}').\mathbf{q}_1'}\right) \left(\frac{(\mathbf{p}' \times \mathbf{q}_3').\mathbf{q}_1'}{(\mathbf{p}' \times \mathbf{q}_3').\mathbf{q}_2'}\right) \tag{12}$$

If $n = 6$ then (12) yields three independent equations on replacing $\mathbf{q} \leftrightarrow \mathbf{q}'$ by each of $\mathbf{q}_4 \leftrightarrow \mathbf{q}_4'$, $\mathbf{q}_5 \leftrightarrow \mathbf{q}_5'$ and $\mathbf{q}_6 \leftrightarrow \mathbf{q}_6'$ in turn. Equation (12) has the general form $\mathbf{a}_q.\mathbf{p}' = 0$ where $\mathbf{a}_q$ is a vector linear in $\mathbf{p}$ that depends on $\mathbf{q} \leftrightarrow \mathbf{q}'$. The three equations $\mathbf{a}_4.\mathbf{p}' = 0$, $\mathbf{a}_5.\mathbf{p}' = 0$ and $\mathbf{a}_6.\mathbf{p}' = 0$ constrain $\mathbf{p}$ to lie on the cubic plane curve $(\mathbf{a}_4 \times \mathbf{a}_5).\mathbf{a}_6 = 0$.

The image correspondence $\mathbf{q}_6 \leftrightarrow \mathbf{q}_6'$ is replaced by a new image correspondence $\mathbf{q}_7 \leftrightarrow \mathbf{q}_7'$ and a second cubic constraint on $\mathbf{p}$ is generated. The two cubic plane curves intersect in nine points but only three of these intersections yield epipoles $\mathbf{p}$ such that (10) holds. The remaining six intersections do not yield possible epipoles. The six intersections include the points $\mathbf{q}_i$ for $1 \leq i \leq 5$.

The advantage of Sturm's method is its elegant mathematical form: it gives closed form solutions for the epipoles. It is also possible to find by an exact algorithm a least squares solution if many cubic plane curves are available. These two approaches have been implemented in MAPLE.

However, because of the algebraic manipulations that are involved, both approaches turned out to be very sensitive to pixel noise. Two methods for reducing the noise sensitivity have been tried. Firstly, the number of manipulations has been reduced by working numerically using only cross ratios and the equations (12). Secondly, the uncertainties in the positions of the image points have been taken into account. In a first implementation without taking uncertainties into account the criterion was very sensitive to pixel noise: in some examples, 0.1 pixel of noise drastically changed the positions of the epipoles. Using N correspondances, partitioned in subsets of four correspondances, the idea is to minimise the criterion

$$C(\mathbf{p}, \mathbf{p}') = \sum_{i=1}^{N/4} \frac{(\tau_i - \tau_i')^2}{\sigma_{\tau_i}^2 + \sigma_{\tau_i'}^2}$$

where τ_i is the cross ratio of the lines $\langle \mathbf{p}, \mathbf{q}_{i_j} \rangle$, $1 \leq j \leq 4$, given by the formula (11), τ_i' is the same with primes, and where $\sigma_{\tau_i}^2$ and $\sigma_{\tau_i'}^2$ are the first order variances on τ_i and τ_i'. The notation $\mathbf{q}_{i_j}$ for $1 \leq j \leq 4$ indicates a subsequence of the $\mathbf{q}_i$. If the noise distribution is the same for all image points $\mathbf{q}_i$, $\mathbf{q}_i'$ then $\sigma_{\tau_i}^2 = \sigma_{pixel}^2 \|\mathbf{grad}(\tau_i)\|^2$, where $\mathbf{grad}(\tau_i)$ is an eight dimensional gradient computed with respect to the $\mathbf{q}_{ij}$ for $1 \leq j \leq 4$. The effect of using the uncertainty in the criterion is that pairs of cross-ratios with large variances will contribute little, whereas others will contribute more.

The problem is that non-linear minimisation techniques are needed. The results of non-linear minimisation are often very dependent on the starting point. Another difficulty is that the position of the minimum is quite sensitive to noise, as will be seen in the experimental section below.

3.2 The Fundamental Matrix Method

The fundamental matrix $\mathbf{F}$ is a generalization of the essential matrix described in [8]. For a given point $\mathbf{m}$ in the first image, the corresponding epipolar line e_m in the second image is linearly related to the projective representation of $\mathbf{m}$. The 3×3 matrix $\mathbf{F}$ describes this correspondence. The projective representation $\mathbf{e}_m$ of the epipolar line e_m is given by

$$\mathbf{e}_m = \mathbf{F}\mathbf{m}$$

Since the point $\mathbf{m}$ corresponding to $\mathbf{m}$ belongs to the line e_m by definition, it follows that

$$\mathbf{m}'^T\mathbf{F}\mathbf{m} = 0 \tag{13}$$

If the image is formed by projection onto the unit sphere then $\mathbf{F}$ is the product of an orthogonal matrix and an antisymmetric matrix. It is then an essential matrix and (13) is the so-called Longuet-Higgins equation in motion analysis [8]. If the image is formed by a general projection, as described in (1), then $\mathbf{F}$ is of rank two. The matrix $\mathbf{A}$ of intrinsic parameters (1) transforms the image to the image that would have been obtained by projection onto the unit sphere. It follows that $\mathbf{F} = \mathbf{A}^{-1T}\mathbf{E}\mathbf{A}^{-1}$, where $\mathbf{E}$ is an essential matrix. Unlike the essential matrix, which is characterized by the two constraints found by Huang and Faugeras [7] 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 7. The essential matrix $\mathbf{E}$ is subject to two independent polynomial constraints in addition to the constraint $\det(\mathbf{E}) = 0$. If $\mathbf{F}$ is known then it follows from $\mathbf{E} = \mathbf{A}^T\mathbf{F}\mathbf{A}$ that the entries of $\mathbf{A}$ are subject to two independent polynomial constraints inherited from $\mathbf{E}$. These are precisely the Kruppa equations. It has also been shown, using the fundamental matrix, that the Kruppa equations are equivalent to the constraint that the two non-zero singular values of an essential matrix are equal.

The importance of the fundamental matrix has been neglected in the literature, as almost all the work on motion has been done under the assumption that intrinsic parameters are known. But if one wants to proceed only from image measurements, the fundamental matrix is the key concept, as it contains the all the geometrical information relating two different images. To illustrate this, it is shown that the fundamental matrix determines and is determined by the epipolar transformation. The positions of the two epipoles and any three of the correspondences $l \overline{\wedge} l'$ between epipolar lines together determine the epipolar transformation. It follows that the epipolar transformation depends on seven independent parameters. On identifying the equation (13) with the constraint on epipolar lines obtained by making the substitutions (8) in (7), expressions are obtained for the coefficients of $\mathbf{F}$ in terms of the parameters describing the epipoles and the homography:

$$\begin{aligned} F_{11} &= bp_3p'_3 \\ F_{12} &= ap_3p'_3 \\ F_{13} &= -ap_2p'_3 - bp_1p'_3 \end{aligned}$$

$$
\begin{aligned}
F_{21} &= -dp'_3p_3 \\
F_{22} &= -cp'_3p_3 \\
F_{23} &= cp'_3p_2 + dp'_3p_1 \\
F_{31} &= dp'_2p_3 - bp_3p'_1 \\
F_{32} &= cp'_2p_3 - ap_3p'_1 \\
F_{33} &= -cp'_2p_2 - dp'_2p_1 + ap_2p'_1 + bp_1p'_1
\end{aligned} \tag{14}
$$

From these relations, it is easy to see that $\mathbf{F}$ is defined only up to a scale factor. Let $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ be the columns of F. It follows from (14) that $p_1\mathbf{c}_1 + p_2\mathbf{c}_2 + p_3\mathbf{c}_3 = 0$. The rank of F is thus at most two. The equations (14), yield the epipolar transformation as a function of the fundamental matrix:

$$
\begin{aligned}
a &= F_{12} \\
b &= F_{11} \\
c &= -F_{22} \\
d &= -F_{21} \\
p_1 &= \frac{F_{23}F_{12} - F_{22}F_{13}}{F_{22}F_{11} - F_{21}F_{12}}p_3 \\
p_2 &= \frac{F_{11}F_{23} - F_{13}F_{21}}{F_{22}F_{11} - F_{21}F_{12}}p_3 \\
p'_1 &= \frac{F_{32}F_{21} - F_{22}F_{31}}{F_{22}F_{11} - F_{21}F_{12}}p'_3 \\
p'_2 &= \frac{F_{11}F_{32} - F_{31}F_{12}}{F_{22}F_{11} - F_{21}F_{12}}p'_3
\end{aligned} \tag{15}
$$

The determinant of the homography is $F_{22}F_{11} - F_{21}F_{12}$. In the case of finite epipoles, it is not null.

A first method to estimate the fundamental matrix takes advantage of the fact that equation (13) is linear and homogeneous in the nine unknown coefficients of $\mathbf{F}$. Thus if eight matches are given then in general $\mathbf{F}$ is determined up to a scale factor. In practice, many more than eight matches are given. A linear least squares method is then used to solve for $\mathbf{F}$. As there is no guarantee, when noise is present, that the matrix $\mathbf{F}$ obtained is exactly a fundamental one, the formulas (15) can not be used, and $\mathbf{p}$ has to be determined by solving the following classical constrained minimization problem

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

This yields $\mathbf{p}$ as the unit norm eigenvector of the matrix $\mathbf{F}^\top\mathbf{F}$ with the smallest eigenvalue. The same processing applies in reverse to the computation of the epipole $\mathbf{p}'$. In contrast with the Sturm method, this method requires only linear operations. It is therefore more efficient and it has no initialization problem.

However the minimum turns out to be sensitive to noise, particularly when the epipoles are far from the centre of the image. Experiments show that this problem is reduced by using the following criterion for minimization:

$$
\min\{d(\mathbf{m}'^T, \mathbf{F}\mathbf{m})^2 + d(\mathbf{m}^T, \mathbf{F}^T\mathbf{m}')^2\} \tag{16}
$$

where d is a distance in the image plane. The criterion has a better physical significance in terms of image quantities. It is necessary to minimize on $\mathbf{F}$ and on $\mathbf{F}^T$ simultaneously

to avoid discrepancies in the epipolar geometry. In doing this non-linear minimization successfully two constraints are important:

- The solution must be of rank two, as all fundamental matrices have this property. Rather than performing a constrained minimization with the cubic constraint $\det(\mathbf{F}) = 0$, it is possible to use, almost without loss of generality, the following representation for $\mathbf{F}$ proposed by Luc Robert:

$$\mathbf{F} = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7x_1 + x_8x_4 & x_7x_2 + x_8x_5 & x_7x_3 + x_8x_6 \end{pmatrix}$$

  One unknown is eliminated directly, and then $\mathbf{F}$ is found by an unconstrained minimization.
- The matrix $\mathbf{F}$ is defined only up to a scale factor. In order to avoid the trivial solution $\mathbf{F} = 0$ at which the minimization routines fail because the derivatives become meaningless, one of the first six elements of $\mathbf{F}$ is normalised by giving it a fixed finite value. However, as the minimization is non-linear convergence results can differ, depending on the element chosen. This feature can be used to escape from bad local minima during minimization.

This second method for computing the fundamental matrix is more complicated, as it involves non-linear minimizations. However, it yields more precise results and allows the direct use of the formulas (15) to obtain the epipolar transformation.

4 Solving Kruppa's Equations: the Continuation Method

Symbolic methods for solving Kruppa's equations are described in [9]. These methods are very sensitive to noise: even ordinary machine precision is not sufficient. Also they require rational numbers rather than real numbers. In this section Kruppa's equations are solved by an alternative method which is suitable for real world use. The current implementation is as follows,

- Do 3 displacements. For each displacement:
  1. Find point matches between the two images
  2. Compute the epipoles
  3. Compute the homography of epipolar lines
  4. Compute the two Kruppa equations
- Solve the six Kruppa equations using the continuation method
- Compute the intrinsic parameters

Three displacements yield six equations in the entries of the matrix $\mathbf{D}$ defined in Sect. 2.1. The equations are homogeneous so the solution for $\mathbf{D}$ 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. The remaining equation is used to discard the spurious solutions. In addition to the six equations, the entries of $\mathbf{D}$ satisfy certain inequalities that are discussed later. These are also useful for ruling out spurious solutions. The problem is that solving a polynomial system by providing an initial guess and using an iterative numerical method will not generally give all the solutions: many of the start 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, as there is only one good 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 to [13] for a 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.

Start System. There are three criteria that guide the choice of a start system: all of its solutions must be known, each solution must be non-singular, and the system must have the same homogeneous structure as the target system. The use of m-homogeneous systems reduces the computational load by eliminating some solutions at infinity, so it is useful to homogenize, but only inhomogeneous systems are discussed here for the sake of simplicity. Thus an acceptable start system is: $x_j^{d_j} - 1 = 0$ for $1 \leq j \leq n$ where n is the number of equations and d_j is the degree of the equation j of the target system. Each equation yields d_j distinct solutions for x_j, and the entire set of $\prod_{j=1}^{n} d_j$ solutions are found by taking all possible combinations of these.

Homotopy. The requirement for the choice of the homotopy (the schedule for transforming the start system into the target system) is that as the transformation proceeds there should be a constant number of solutions which trace out smooth paths and which are always nonsingular until the target system is reached. It has been shown by years of practice that the following homotopy suffices:

$$H(\mathbf{x}, t) = (1 - t)e^{i\theta}G(\mathbf{x}) + tF(\mathbf{x})$$

where $G(\mathbf{x})$ is the start system, and $F(\mathbf{x})$ is the target system.

Path tracking. Path tracking is the process of following the solutions of $H(\mathbf{x}, t) = 0$ as t is increased from 0 to 1. These solutions form d ***continuation paths***, where d is the Bezout number of the system, characterising the number of solutions. To track a path from a known solution $(\mathbf{x}^0, t^0)$, the solution is predicted for $t = t^0 + \Delta t$, using a Taylor expansion for H, to yield $\Delta\mathbf{x} = -J_x^{-1} J_t \Delta t$, where J_x and J_t are the Jacobians of H with respect to $\mathbf{x}$ and t. The prediction is then corrected using Newton's method with t fixed at the new value to give corrections steps of $\Delta\mathbf{x} = -J_x^{-1} H(\mathbf{x}, t)$.

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 work. 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 one minute for our problem.

5 Computing the Intrinsic Parameters

In this section the relation between the image of the absolute conic and the intrinsic parameters is given in detail. The most general matrix $\mathbf{A}$ occurring in (1) can be written:

$$\mathbf{A} = \begin{bmatrix} -fk_u & fk_u \cot(\theta) & u_0 \\ 0 & -fk_v \operatorname{cosec}(\theta) & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

- k_u, k_v, are the horizontal and vertical scale factors whose inverses characterize the size of the pixel in world coordinates units.
- u_0 and v_0 are the image center coordinates, resulting from the intersection between the optical axis and the image plane.
- f is the focal length
- θ is the angle between the directions of retinal axes. This parameter is introduced to account for the fact that the pixel grid may not be exactly orthogonal. In practice θ is very close to $\pi/2$.

As f cannot be separated from k_u and k_v it is convenient to define products $\alpha_u = -fk_u$ and $\alpha_v = -fk_v$. This gives five intrinsic parameters. This is exactly the number of independent coefficients for the image w of the absolute conic thus the intrinsic parameters can be obtained from w. The equation of w is [3]:

$$\mathbf{y}^T \mathbf{A}^{-1T} \mathbf{A}^{-1} \mathbf{y} = 0$$

It follows that $\mathbf{D} = \mathbf{A}\mathbf{A}^T$. Up to a scale factor the entries δ_{ij} and δ_i of $\mathbf{D}$ are related to the intrinsic parameters by

$$\begin{aligned} \delta_1 &= v_0 \\ \delta_2 &= u_0 \\ \delta_3 &= u_0 v_0 - \alpha_u \alpha_v \cot(\theta) \operatorname{cosec}(\theta) \\ \delta_{12} &= -1 \\ \delta_{23} &= -u_0^2 - \alpha_u^2 \operatorname{cosec}^2(\theta) \\ \delta_{13} &= -v_0^2 - \alpha_v^2 \operatorname{cosec}^2(\theta) \end{aligned}$$

From these relations it is easy to see that the intrinsic parameters can be uniquely determined from the Kruppa coefficients, provided the five following conditions hold:

$$\begin{aligned} \delta_{13}\delta_{12} &> 0 \\ \delta_{23}\delta_{12} &> 0 \\ \delta_{13}\delta_{12} - \delta_1^2 &> 0 \\ \delta_{23}\delta_{12} - \delta_2^2 &> 0 \\ \frac{(\delta_3\delta_{12} + \delta_1\delta_2)^2}{(\delta_{13}\delta_{12} - \delta_1^2)(\delta_{23}\delta_{12} - \delta_2^2)} &\leq 1 \end{aligned} \tag{18}$$

If one of the conditions (18) doesn't hold then there is no physically acceptable calibration compatible with the Kruppa coefficients δ_{ij} and δ_i. This is a strong condition which rules out many spurious solutions obtained by solving five of the Kruppa equations. It is interesting to note that if a four-parameter model is used with $\theta = \pi/2$ then there is the additional constraint $\delta_3 = -\delta_1\delta_2/\delta_{12}$ which replaces the last one of (18). It can be also verified that the calibration parameters depend only on the ratios of Kruppa coefficients, so that the scaling of them doesn't modify their value, as expected.

6 Experimental Results

The results of experiments with computer generated data are described. The coordinates of the projections of 3D points are computed using a realistic field of view and realistic values for the extrinsic and intrinsic parameters. For each displacement 20 point matches are selected and noise is added.

6.1 Computation of the epipoles

The results for the determination of the epipoles in the first image are presented. The values obtained by the two algorithms (the Sturm method based on weighted cross-ratios, and the fundamental matrix method) are given, as well as the relative error with respect to the exact solution. The results in the second image are always similar to those in the first image.

pixel noise	motion 1 [.497578 .01443363 .49306] [-335.50 985.39 325.14]				motion 2 [0 .05 0] [0 0 400]				motion 3 [0.1 0 0] [50 20 20]			
	Sturm		Fund. Matrix		Sturm		Fund. Matrix		Sturm		Fund. Matrix	
0	-414.42	3115.64	-414.42	3115.64	246.09	255.64	246.09	255.64	1846.40	1199.34	1846.40	1199.34
0.01	-413.96	3113.82	-414.53	3115.57	246.06	255.66	246.10	255.57	1847.89	1200.50	1841.39	1195.77
	0.1	0.06	0.02	0.02	0.01	0.007	0.04	0.02	0.08	0.1	0.27	0.29
0.1	-410.05	3098.17	-415.47	3114.49	245.69	255.70	246.11	255.00	1864.07	1212.39	1792.52	1161.69
	1	0.5	0.2	0.03	0.16	0.02	0.008	0.25	1	1.1	2.9	3.1
0.2	-406.08	3082.27	-416.27	3112.29	245.58	255.90	246.05	254.47	1889.58	1229.63	1731.60	1120.51
	2	1.1	0.4	0.1	0.2	0.1	0.016	0.45	2.3	2.5	6.2	6.2
0.5	-396.32	3043.11	-417.13	3099.29	244.45	256.36	245.40	253.64	2045.52	1325.10	1527.21	988.02
	4.3	2.3	0.6	0.5	0.6	0.3	0.3	0.8	10.7	10.5	17	17
1.0	-386.10	3001.28	-413.24	3055.80	239.82	256.12	242.62	254.90	762.82	554.48	1201.05	785.79
	6.8	3.6	0.2	1.9	2.5	0.2	1.4	0.3	58	53	35	34
2.0	-333.19	2772.54	-385.41	2889.29	226.05	284.41	230.44	267.73	801.64	592.33	785.28	536.30
	19	11	7	7.2	8	11	6.3	4.7	56	50	57	55

From these results, it can be seen that the fundamental matrix method is more robust. It is also computationally very efficient since it involves only a linear least squares minimisation and a 3×3 eigenvector computation. A second point worth noting is that the stability of the position of the epipole depends strongly on the displacement that is chosen.

Other experiments not reported here due to lack of space show that if more matches are available then the precision of the determination of the epipoles can be improved.

6.2 Intrinsic parameters

The intrinsic parameters that have been computed using two displacement sequences are presented. The first sequence consists of *motion 1*, *motion 4*, *motion 2*. The second sequence consists of *motion 1*, *motion 4*, *motion 3*.

	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
0 pixels	640.125	943.695	246.096	255.648	0
0.01 pixels	597.355	940.403	248.922	259.196	0.02
	6.68	0.34	1.14	1.38	
0.1 pixels	520.126	904.744	275.120	280.601	0.09
	18.7	4.1	11.8	9.7	
0.2 pixels	175.204	867.214	565.234	291.162	0.4
	72.6	8.1	129.6	13.8	

	α_u	α_v	u_0	v_0	$\theta - \frac{\pi}{2}$
0.01 pixels	699.815	948.106	174.723	245.112	0.018
	9.3	0.46	29.0	4.12	
0.1 pixels	687.814	989.538	149.055	257.030	0.004
	7.4	4.85	39.4	0.54	
0.2 pixels	552.601	993.837	269.278	283.458	0.013
	13.6	5.31	9.41	10.87	
0.5 pixels	433.894	957.018	358.904	308.191	0.13
	32.2	1.4	45	20.5	
1.0 pixel	477.043	724.909	137.052	258.763	0.3
	25.4	23.2	44.3	1.2	

More precise results are obtained if more than three camera displacements are available.

These results demonstrate the feasibility of the method in real environments, provided image points can be located with a sufficient precision. This precision is already achievable using special patterns.

7 Conclusion and Perspectives

A method for the on-line calibration of the intrinsic parameters of a camera has been described. The method is based on the estimation of the epipolar transformations associated with camera displacement. Three epipolar transformations arising from three different displacements are sufficient to determine the camera calibration uniquely. The epipolar transformations can in principle be obtained by tracking a number of salient image points while the camera is moving. It is therefore not necessary to interrupt the action of the vision system in order to point the camera at a special test pattern.

The feasibility of the method is demonstrated by a complete implementation which is capable of finding the intrinsic parameters provided a sufficient number of point matches are available with a sufficient precision. However, the precision required to obtain acceptable calibrations is at the limit of the state-of-art feature detectors.

The next step is thus to find efficient methods to combat noise. The key idea is to compute the uncertainty explicitly. The results have shown that some displacements yield epipolar transformations that are very sensitive to pixel noise, whereas, some yield transformations that are more robust. Methods for characterising "bad" displacements are currently being investigated. In particular, it has been shown that pure translations lead to degenerate cases, thus yielding results that are very sensitive to noise. However, it is not sufficient to know *a priori* which displacements are best because as the camera is not yet calibrated they cannot be applied. If the uncertainty in the epipolar transformation obtained from a given displacement can be computed it can be the basis of a decision whether to use the transformation for the computations, to discard it and use another one, or to take it into account only weakly. The final aim is to obtain acceptable calibrations using real images.

References

1. R. Deriche and G. Giraudon. Accurate corner detection: An analytical study. In *Proceedings ICCV*, 1990.
2. Rachid Deriche and Olivier D. Faugeras. Tracking Line Segments. *Image and vision computing*, 8(4):261–270, November 1990. A shorter version appeared in the Proceedings of the 1st ECCV.

3. O.D. Faugeras. *Three-dimensional computer vision.* MIT Press, 1992. To appear.
4. O.D. Faugeras and G. Toscani. The calibration problem for stereo. In *Proceedings of CVPR'86*, pages 15–20, 1986.
5. C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. 4th Alvey Vision Conf.*, pages 189–192, 1988.
6. O. Hesse. Die cubische Gleichung, von welcher die Lösung des Problems der Homographie von M. Chasles abhängt. *J. reine angew. Math.*, 62:188–192, 1863.
7. T.S. Huang and O.D. Faugeras. Some properties of the e matrix in two view motion estimation. *IEEE Proc. Pattern Analysis and Machine Intelligence*, 11:1310–1312, 1989.
8. H.C. Longuet-Higgins. A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature*, 293:133–135, 1981.
9. S.J. Maybank and O.D. Faugeras. A Theory of Self-Calibration of a Moving Camera. *The International Journal of Computer Vision*, 1992. Submitted.
10. Rudolf Sturm. Das Problem der Projektivität und seine Anwendung auf die Flächen zweiten Grades. *Math. Ann.*, 1:533–574, 1869.
11. G. Toscani, R. Vaillant, R. Deriche, and O.D. Faugeras. Stereo camera calibration using the environment. In *Proceedings of the 6th Scandinavian conference on image analysis*, pages 953–960, 1989.
12. Roger Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. In *Proceedings CVPR '86, Miami Beach, Florida*, pages 364–374. IEEE, June 1986.
13. 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.

This article was processed using the LaTeX macro package with ECCV92 style