Plane-based Calibration of a Camera with Varying Focal Length: the Centre Line Constraint

Pierre GURDJOS and René PAYRISSAT
IRIT-UPS, TCI
118, route de Narbonne
31062 Toulouse Cedex 4, FRANCE
Pierre.Gurdjos@irit.fr Rene.Payrissat@irit.fr

Abstract

This paper deals with the problem of calibrating a (moving) camera with varying focal length, from n views of a planar pattern with a known Euclidean structure. The main issue under discussion is to find a new method whose complexity does not dramatically increase with the number n of views, contrary to existing methods. Our contribution is to relate this calibration problem to the Centre Line (*CL*) constraint, that is the principal point locus when planar figures are in perpective correspondence, in accordance with Poncelet's theorem. We demonstrate that the *CL* equation is irrespective of the focal length and holds for each view, with only three unknown parameters whose values are constant in the images. We define its analytic equation with coefficients computed from the world-plane to image homography matrices. An important aspect is that we can make use of it as a linear cost function that expresses a geometric error (instead of algebraic errors in existing methods). We explain why an "optimal" solution can be obtained when pixels are rectangular. The simulations on synthetic data and an application with real images confirm the two strong points of our method with respect to existing ones: a lower computation cost and a better system conditioning that permits to obtain more accurate results.

1 Introduction

At the highest level of flexibility, the "autocalibration of a projective camera" problem consists in recovering the camera metric information from non-metric information and self-consistency constraints (read [9] or [1] for historical background). Fundamentally, one very determining result has been to highlight some imaginary projective entities, like the absolute conic or the dual absolute quadric, enabling to define a generic framework based on properties of the direction bases. In particular, the problem of "autocalibration from planar scenes" [8][4], *i.e.* from multiple views of at least one reference plane, can be stated within this framework.

Specifically, the issue at stake here is about "plane-based calibration" [12][7][3] that can be seen as a specialization of the "autocalibration from planar scenes" problem. By definition, plane-based calibration requires metric information about a reference plane Π (*e.g.* coordinate system, angle, length ratio, etc.). An important advantage of plane-based calibration is that the estimation of Π-plane to image homographies are very stable and accurate. Plane-based calibration includes the case of a "single view calibration from multiple planes" where some knowledge about the camera internal parameters (*e.g.* zero-skew and known aspect ratio) are assumed, providing that three mutually conjugated vanishing points are known [2] or in presence of parallelepipeds [10]. In a seminal paper [12] about plane-based calibration, Z. Zhang considers the same constraints than those used in autocalibration from planar scenes, but substitutes the world to image homographies (supposed to be known up to a similarity) induced by a plane Π for the inter-image homographies. For varying intrinsic parameters, Sturm and Maybank [7] increase the scope of this method by showing that coefficients depending on these varying parameters can be considered as additional unknowns of the problem. In some ways, these plane-based calibration algorithms offer a compromise solution between the high accuracy of photogrammetric methods and the flexibility of autocalibration. These algorithms are easy to implement and fast, because the solution is obtained by solving linear equation systems to least squares by non iterative procedures. They also prove to be efficient, providing that there are no critical camera displacements [7]. The snag with these algorithms is that they minimize algebraic distances and a good choice of "system normalization" (as we will discussed later) is often necessary to obtain reliable results. When only the focal length can vary, an optimal solution (attaining the Cramer-Rao lower bound) is proposed in [5]. Nevertheless, it requires an off-line pre-calibration step (not described) and Newton iterations in order to only consider the different unknown focal lengths uncalibrated.

Motivation. Within the framework of plane-based calibration of a camera with varying internal parameters, our aim is to find a new algorithm whose complexity does not dramatically increase with the number of views, contrary to the existing method [7]. In the case of a varying focal length, we suggest to split plane-based calibration problem in two. The first subproblem is that of recovering the three constant internal parameters (aspect ratio and principal point coordinates). The second subproblem is then to compute the focal length for each partially calibrated camera. We relate the first sub-problem to the properties of the Centre Line (*CL*) that is the principal point locus when planar figures are in perspective correspondence. We show that the geometric properties of the Centre Line allows us to define an "optimal" cost function based on orthogonal distances in the images. For the two subproblems, this geometric formulation helps us intuitively to avoid the singularities, already exhaustively listed in [7]. Simulations show the very good performance of the *CL*-based algorithm *vs* Sturm & Maybank's algorithm (regarding both complexity and accuracy).

Notations. A point or line is denoted by a $\mathsf{sans\ serif}$ letter p, L,... and the vector that represents it by the same lowercase bold letter $\mathbf{p}, \mathbf{l}$,... A matrix is denoted by an uppercase bold letter $\mathbf{M}$,... or a Greek bold letter $\boldsymbol{\omega}$,.... We use $\mathrm{diag}(d_1, \cdots, d_n)$ for denoting the diagonal matrix of order n with the elements $d_1, \cdots, d_n$ on the main diagonal . The column vectors of matrix $\mathbf{M}$ are denoted by $\mathbf{m}_j$ and the elements of $\mathbf{m}_j$ are denoted by M_{ij}. A set of n matrices of the form $\mathbf{M}$ is denoted by $\mathbf{M}^{(1)}, \mathbf{M}^{(2)} ... \mathbf{M}^{(n)}$. A scalar is denoted by a non-bold letter X, u or α. Notation $\mathbf{x} \sim \mathbf{y}$ means that $\mathbf{x}$ is equal to $\mathbf{y}$ up to a scalar factor while notation $a \approx b$ means that "a approximately equal to b".

2 Plane-based calibration constraints

Projective Camera model. The Euclidean coordinates (X, Y, Z) of a point and its imaged coordinates (u, v) are related by the projection equation $[u, v, 1]^\top \sim \mathbf{P}[X, Y, Z, 1]^\top$. The perspective projection 3×4 matrix $\mathbf{P}$ can be decomposed as $\mathbf{P} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}]$ where $\mathbf{A}$ is the 3×3 upper triangular matrix of the camera internal parameters and $(\mathbf{R}, \mathbf{t})$ is a 3D rigid displacement. We will use the following decomposition $\mathbf{A} = \mathbf{A}_c\mathbf{A}_f$, by writing:

$$\mathbf{A} = \begin{bmatrix} f & \sigma f & u_0 \\ 0 & \kappa f & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathbf{A}_c = \begin{bmatrix} 1 & \sigma & u_0 \\ 0 & \kappa & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \ \mathbf{A}_f = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where σ is a skewness factor, f is the focal length in pixels, κ is an aspect ratio parameter (unit-less) and (u_0, v_0) are the pixel coordinates of the principal point. Pixels are very often supposed to be rectangular in the literature so, all along the paper, we will assume $\sigma = 0$ (hypothesis known as the zero-skew constraint).

Let $\mathcal{P}(\mathbb{R}^3)$ be the 3-dimensional projective space and $\mathcal{B}$ a basis of $\mathcal{P}(\mathbb{R}^3)$ such that the equation of the plane at infinity Π_∞ is $T = 0$. The absolute conic Ω has equation $X^2 + Y^2 + Z^2 = T = 0$ with respect to basis $\mathcal{B}$. The matrix $\boldsymbol{\omega} = \mathbf{A}^{-\top}\mathbf{A}^{-1}$ represents the image of the absolute conic Ω under $\mathbf{P}$. Hence the recovery of $\boldsymbol{\omega}$ is closely connected to the intrinsic calibration of the camera (*i.e.* the recovery of matrix $\mathbf{A}$).

Internal parameter constraints. The image of the absolute conic Ω under $\mathbf{P}$ is represented by:

$$\boldsymbol{\omega} \sim \begin{bmatrix} \kappa^2 & 0 & -\kappa^2 u_0 \\ 0 & 1 & -v_0 \\ -\kappa^2 u_0 & -v_0 & \kappa^2 u_0^2 + v_0^2 + \kappa^2 f^2 \end{bmatrix}. \quad (2)$$

Let $\omega_{ij} (i \leq j)$ be the elements of $\boldsymbol{\omega}$. Under the zero-skew assumption, the following four equations can be used as autocalibration constraints [1]:

$$\omega_{12} = 0, \qquad \omega_{11} = \kappa^2 \omega_{22}, \quad (3)$$

$$u_0\omega_{11} = -\omega_{13}, \qquad v_0\omega_{22} = -\omega_{23}. \quad (4)$$

Circular points constraints. Let us consider the case of a plane Π with equation $Z = 0$ and adjoin to it two conjugate complex points $\mathrm{c}_\pm$ on the line at infinity ($Z = T = 0$), known as the *circular points*. These points have coordinate vectors $\mathbf{c}_\pm = [1, \pm i, 0]^\top$, where $i^2 = -1$. Let $\mathbf{H}$ be the world-to-image 3×3 homography matrix induced by Π. Circular points verify $X^2 + Y^2 = 0$ and therefore lie on the absolute conic Ω. Their images under $\mathbf{H}$ lie on the image of Ω, which means that $(\mathbf{H}\mathbf{c}_\pm)^\top \boldsymbol{\omega}(\mathbf{H}\mathbf{c}_\pm) = 0$. Requiring that both real and imaginary parts of this equation be zero yields the two linear equations:

$$\mathbf{h}_1^\top \boldsymbol{\omega} \mathbf{h}_2 = 0, \quad \mathbf{h}_1^\top \boldsymbol{\omega} \mathbf{h}_1 = \mathbf{h}_2^\top \boldsymbol{\omega} \mathbf{h}_2, \quad (5)$$

where $\mathbf{h}_1$ and $\mathbf{h}_2$ are the first two columns of $\mathbf{H}$. These constraints are linear in the elements of $\boldsymbol{\omega}$ and form the basis equations of the plane-based calibration problem.

[1] These constraints are derived from the form of $\boldsymbol{\omega}$ in (2) but, to our knowledge, the constraints (4) have never been subjected to a geometrical interpretation. We suggest that they arise from a pole-polar property. In the projective plane, the line at infinity has a pole p_0 with respect to any proper conic Σ and this point is the centre of Σ. In particular, for the absolute conic, we can write $\Omega\mathbf{p}_0 \sim \mathbf{l}_\infty$ where the vector $\mathbf{p}_0 = [0, 0, 1]^\top$ represents the centre of Σ and $\mathbf{l}_\infty = [0, 0, 1]^\top$ represents the line at infinity. This matrix equation is transformed under $\mathbf{A}$ into $\boldsymbol{\omega}[u_0, v_0, 1]^\top = \mathbf{l}_\infty$, since $\mathbf{l}_\infty$ is invariant under any affine transformation. The transformed matrix equation yields only two independent equations that are strictly equivalent to (4).

2.1 Related works

To our knowledge, there is no non iterative plane-based calibration method other than those based on equations (5) and discussed in the introduction.

A. Constant internal parameters. By assuming constant internal parameters, Zhang [12] gives a closed-form solution to the calibration problem : with n images yielding n homography matrices $\mathbf{H}^{(k)}$, there are $2n$ equations taken from (5) while the unknown symmetric matrix $\boldsymbol{\omega}$ has 5 degrees of freedom. Consequently at least 3 images are required to recover $\boldsymbol{\omega}$. Under the zero-skew assumption, 2 images suffice.

B. Varying internal parameters. In [7], Sturm & Maybank extend[2] Zhang's work [12] to the case of varying internal parameters. In their paper, pixels are supposed to be rectangular ($\sigma = 0$). Two cases are considered : [i] only the focal length f vary or [ii] f varies in conjunction with the principal point. These varying parameters are considered as additional unknowns of the basic plane-based constraints (5). Typically, for n images, in the case [i] (resp. [ii]), the system will have $2n$ equations with $3 + n$ (resp. $5 + n$) unknowns.

2.2 Discussion

From a computational point of view, the solution to the problem 2.1.B seems to us to be questionable, at least for two reasons. Basically, our criticisms apply to the algorithm complexity and system conditioning.

Complexity. Regarding algorithms for plane-based calibration (including the varying focal length case) from n images, the problem is to solve a set of homogeneous equations of the form $\mathbf{Mx} = \mathbf{0}$, where $\mathbf{x}$ is the unknown m-vector with $(m-1)$ degrees of freedom and $\mathbf{M}$ is the $2n \times m$ matrix of coefficients computed from matrices $\mathbf{H}^{(k)}$ in (5). When the system is over-determined (*i.e.* $2n > m$), the task it to find the vector $\mathbf{x}$ that minimizes the cost function $||\mathbf{Mx}||^2$ under the constraint $||\mathbf{x}|| = 1$. Such a least-squares solution requires the singular value decomposition SVD of $\mathbf{M}$ (or equivalently the computation of the spectral decomposition of $\mathbf{M}^\top\mathbf{M}$). With regard to problem 2.1.B, in the Sturm & Maybank's calibration algorithm for varying focal length, this computation applies to a matrix whose dimension is $2n \times (4 + n)$, where n is the number of views. Such computation is extensive and does not suit to real time-like processing of video image sequences. Indeed, when solving a least-squares problem[3] from n images, the total number of flops for carrying out the SVD of $\mathbf{M}$ is $16n^3 + 160n^2 + 512n + 512$ (computed from [1], pp.558). If we consider, for instance, a sequence of $n = 100$ images, the matrix $\mathbf{M}$ will have 200 rows and 104 columns and the minimization will require 17.6 Megaflops.

Conditioning. The cost functions used in [12][7] are based on algebraic distances. As a general rule, it is known that such quantities are not geometrically or statistically meaningful and the obtained solutions may not be those expected intuitively [1]. Although we do not provide the proof here, we also claim that, in the case of a varying focal length, input homographies "corrupted with the same amount of noise" may contribute differently to the estimation problem depending on the focal length value, which leads to an undesirable row weighting. Moreover, the elements of matrix $\mathbf{M}$ and vector $\mathbf{x}$ may differ in magnitude in a very sizeable way. That leads to a bad system conditioning that Sturm and

[2] Although their paper does not refer to Zhang's work.

[3] Only the matrices $\Sigma, \mathbf{V}$, appearing in the SVD decomposition $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top$, are computed.

Maybank improve by column rescaling such as to have equal norms [7]. The least we can say is that, according to what we have noticed when running simulations, this rescaling is not optional *at all*.

In the following, we want to define a linear cost function that does not reveal such disadvantages. It can be achieved if we replace the algebraic constraints on the imaged absolute conic by geometric constraints, deduced from two planar figures in perspective correspondence, defining a pencil of "centre lines" irrespective of the focal length.

3 The Centre Line-based constraint

3.1 Geometrical background

The geometric properties deduced from two planar figures in perspective correspondence and, in particular, about the positions of the centres of projection, is known for a long time, notably thanks to the works of J.V. Poncelet in the middle of the XIX century [6]. Fundamentally, Poncelet demonstrated the following theorem in a a purely geometric way: *"When a planar figure is a central projection of another planar figure, these figures remain in perspective when one rotates the plane of the first around its intersection with the plane of the second, and the centre of projection describes a circle* C*, lying on a plane which is perpendicular to this intersection".* If we consider the image of a plane Π, this leads us to the following definition.

Definition 1 (Centre line) *The centre line* L_C *is the line segment obtained by the orthogonal projection of the circle* C *onto the image plane.*

Property 2 *According to Poncelet's theorem, the centre line* L_C *is the geometric locus of the principal point.* L_C *is orthogonal to the vanishing line* L_∞^Π *of* Π *and the orientations of these two lines are invariant to variations of the focal length.*

In the following, we define explicitly the analytic equation of the centre line. We demonstrate that this equation is irrespective of the focal length f and we use it as a constraint in order to recover u_0, v_0, κ, *i.e.* the elements of matrix $\mathbf{A}_c$ in (1). This allows us to integrate a large amount of views in the system since any additional view does not introduce a new unknown. The focal length is then directly computed by solving the partially calibrated system (5).

Proposition 3 *Let* $\mathbf{H}$ *be the world-to-image homography matrix and let* H_{ij} *be the element* (i,j) *of* $\mathbf{H}$*. The vector* $\mathbf{l}_C = [-\varphi_1\kappa^2, -\varphi_2, \varphi_3\kappa^2 + \varphi_4]^\top$*, where:*

$$\begin{array}{ll} \varphi_1 = (H_{31}^2 + H_{32}^2)(H_{31}H_{12} - H_{11}H_{32}), & \varphi_2 = (H_{31}^2 + H_{32}^2)(H_{31}H_{22} - H_{21}H_{32}), \\ \varphi_3 = (H_{31}H_{12} - H_{11}H_{32})(H_{11}H_{31} + H_{12}H_{32}), & \varphi_4 = (H_{31}H_{22} - H_{21}H_{32})(H_{21}H_{31} + H_{22}H_{32}), \end{array} \tag{6}$$

represents the line coordinates of the centre line L_C *with respect to the affine image frame.*

Proof. Let us initially eliminate algebraically ω_{33} from equations (5). We obtain a new equation linear in the terms of $\boldsymbol{\omega}$:

$$\boldsymbol{\varphi}^\top [\omega_{13}, \omega_{23}, \omega_{11}, \omega_{22}]^\top = 0, \tag{7}$$

where $\boldsymbol{\varphi} = [\varphi_1, \varphi_2, \varphi_3, \varphi_4]^\top$ is defined as in (6). Introducing relations (3)(4) in (7), we obtain:

$$-\varphi_1\kappa^2 u_0 - \varphi_2 v_0 + (\varphi_3\kappa^2 + \varphi_4) = 0 \tag{8}$$

Equation (8) tells us that the principal point with pixel coordinates (u_0, v_0) lies on a certain line L, represented by $\mathbf{l} = [-\varphi_1\kappa^2, -\varphi_2, \varphi_3\kappa^2 + \varphi_4]^\top$ in the image frame. We now demonstrate that L coincides with centre line L_C. Since line L contains (u_0, v_0), using property 2, it suffices to show that line L is orthogonal to the vanishing line L_∞^Π of 3D plane Π. It is known that the vector $\mathbf{l}_\infty^\top$ representing line L_∞^Π, is equal (projectively speaking) to the third row of the adjoint matrix of $\mathbf{H}$. Hence, after some computations, we find: $\mathbf{l}_\infty^\top \sim [H_{21}H_{32} - H_{31}H_{22}, H_{31}H_{12} - H_{11}H_{32}, H_{11}H_{22} - H_{12}H_{21}]$. With respect to the (affine) image frame, it is equivalent to show that $\mathbf{l}$ and $\mathbf{l}_\infty$ are orthogonal with respect to $\Omega^* \sim \mathrm{diag}(1/\kappa^2, 1, 0)$. This condition holds since it is straightforward to verify that $\mathbf{l}^\top\Omega^*\mathbf{l}_\infty = 0$. In conclusion, the vector $\mathbf{l}$ represents the coefficients of centre line L_C with respect to the image frame. ■

3.2 A two-step procedure for plane-based calibration

In the previous section, it has been shown that the equation of centre line L_C defined in (8) results in a linear constraint on the unknown 4-vector $\mathbf{x} = \lambda\left[\kappa^2 u_0, v_0, \kappa^2, 1\right]^\top$, $\lambda \neq 0$. Its elements are supposed to be constant in the images, so the number of degrees of freedom in equation (7) is 3, whatever the number n of images. This allows to estimate u_0, v_0, κ in a first step, from $n \geq 3$ homographies. The second step is then to directly recover the focal length for each partially calibrated view. We now describe precisely these two steps.

3.2.1 Recovering the constant internal parameters

With n views providing n world-to-image homography matrices $\mathbf{H}^{(k)}$, the problem is to find $\hat{\mathbf{x}}$ such that:

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{ ||\Phi\mathbf{x}||^2 \right\} \text{ under the constraint } ||\mathbf{x}|| = 1, \text{ where } \Phi = \begin{bmatrix} \boldsymbol{\varphi}^{(1)\top} \\ \vdots \\ \boldsymbol{\varphi}^{(n)\top} \end{bmatrix}. \tag{9}$$

Φ is the $n \times 4$ matrix formed by stacking the row vectors $\boldsymbol{\varphi}^{(k)\top} = \left[\varphi_1^{(k)}, \varphi_2^{(k)}, \varphi_3^{(k)}, \varphi_4^{(k)}\right]$ computed from $\mathbf{H}^{(k)}$, $k \leq 1 \leq n$, like explained in (6). Once $\mathbf{x}$ has been recovered, we can easily compute $\hat{u}_0, \hat{v}_0, \hat{\kappa}$ from equations (3)(4).

Search of a geometric distance. A very interesting aspect of our formulation, using a centre-line approach, is that we can easily transform the cost function $||\Phi\mathbf{x}||^2$ into a sum of squares of Euclidean distances. It is worthy of note that, under the assumption of Gaussian noise, the minimization of a geometric cost function in the images is equivalent to a Maximum Likelihood Estimation (MLE) [1]. It is also known that the geometric description of constraints is quite advantageous because system conditioning (but also the detection of degenerated configurations) becomes clearer. The following proposition enables us transform the cost function into a geometric function based on orthogonal distances in the images.

Proposition 4 *Let $\bar{\Phi}$ the matrix obtained by normalizing each row $\boldsymbol{\varphi}^\top = [\varphi_1, \varphi_2, \varphi_3, \varphi_4]$ as $\bar{\boldsymbol{\varphi}}^\top = \frac{\boldsymbol{\varphi}^\top}{\sqrt{\varphi_1^2 + \varphi_2^2}}$. When $\kappa = 1$, we seek $\hat{\mathbf{x}}$ such that:*

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{ \left\| \bar{\Phi}\mathbf{x} \right\|^2 \right\} = \operatorname*{argmin}_{\mathbf{x}} \left\{ \sum_{j=1}^{n} d^2(\mathsf{m}_0, \mathsf{L}_C^{(j)}) \right\} \tag{10}$$

where $d(\mathsf{m}_0, \mathsf{L}_C)$ *is the Euclidean distance from principal point* m_0 *to centre line* L_C.

This proposition means that the residual error in (10) is the sum of n squared orthogonal distances. In practice, we know that $\kappa \approx 1$, so this proposition is still acceptable.

Complexity-wise, the least-squares problem (10) is then solved via the SVD decomposition $\Phi = \mathbf{U}\Sigma\mathbf{V}^\top$ by computing $\hat{\mathbf{x}} = \mathbf{v}_*$, where $\mathbf{v}_*$ is the column of $\mathbf{V}$ associated with the smallest singular value. Actually, only Σ and $\mathbf{V}$ are computed so the amount of work required by this step of the algorithm is about $69n + 512$ flops (including row normalizations of proposition 4), where n is the number of images.

3.2.2 Recovering the focal lengths

For each view k, let $\mathbf{G} = \mathbf{A}_c^{-1}\mathbf{H}^{(k)}$, where $\mathbf{A}_c$ is constructed from $\hat{u}_0, \hat{v}_0, \hat{\kappa}$ as in (1). The image of the absolute conic now modifies to $\mathbf{A}_c^\top \boldsymbol{\omega}^{(k)} \mathbf{A}_c \sim \mathrm{diag}(1, 1, f^{(k)^2})$. It can be seen that the problem is to solve for $F = f^{(k)^2}$ the matrix equation $\Psi\,[F, 1]^\top = 0$ where:

$$\Psi = \begin{bmatrix} G_{12}^2 + G_{21}^2 - G_{12}^2 - G_{22}^2 & G_{31}^2 - G_{32}^2 \\ G_{11}G_{12} + G_{21}G_{22} & G_{31}G_{32} \end{bmatrix}. \tag{11}$$

Let $\bar{\Psi}$ be the matrix obtained by normalizing the columns Ψ_i of Ψ such that $\|\bar{\Psi}_i\| = 1$ and let $\mathbf{Q} = \bar{\Psi}^\top\bar{\Psi}$. The explicit least-squares solution $\hat{F}$ can be computed directly and we obtain: $\hat{F} = \alpha\frac{\lambda - Q_{11}}{Q_{12}}$, where $\alpha = \frac{\|\Psi_1\|}{\|\Psi_2\|}$, $\lambda = \frac{1}{2}(Q_{11} + Q_{22} - \sqrt{(Q_{11} - Q_{22})^2 + 4Q_{12}^2})$. For each view, the flop count associated with this step of the algorithm is about 60 flops.

3.2.3 Singularities

The singularities of calibration must occur at two levels in our method, *i.e.* when recovering the constant internal parameters (level 1) and when recovering the focal lengths (level 2) . The singularities are basically the same as those described in [7] but our geometrical framework enables us to give some justifications. We describe in the following some of the most significant. At level 1, degeneracies occur if, in two of the three required views, vanishing lines associated with the observed plane Π are parallel. It happens in particular when the camera displacement is either a translation, a rotation with axis parallel or perpendicular to Π or a composition of these displacements. Consequently, the vanishing lines in different views must have different orientations in order to define the principal point as the vertex of the pencil of centre lines. In the majority of cases, adding rotations around the Z-axis can noticeably improve the results. At level 2, regarding system (11), there is no solution for f^2 in two cases: [i] $G_{31} = \pm G_{32}$ [ii] $G_{31} = 0$ or $G_{32} = 0$. These cases correspond to particular orientations of the world coordinate system with respect to the camera coordinate system : they are unnecessary to describe them since they can easily vanish by applying a rotation around the Z-axis (*i.e.* the normal to Π) of the world frame (homographies are estimated up to a similarity). For instance, [ii] holds as soon as the u-axis (resp. v-axis) of the camera frame is parallel to the X-axis (resp. Y-axis) of the world frame (*i.e.* associated vanishing points are points at infinity).

4 Experiments

Simulation[4]**.** We carry out comparison tests about the Centre Line-based method (CL) *vs* Sturm & Maybank's method (SM) for a moving camera with variable focal length. A simulation is run with 8 cameras shooting a calibration object. The calibration object is a planar grid containing $11 \times 11 = 121$ points with dimension $10cm \times 10cm$. The image resolution is 512×512 pixels. All the internal parameters of the camera are supposed to be unknown. The parameters $u_0 = 256$, $v_0 = 256$, $\kappa = 1$ are constant. We divide the set of 8 cameras in 2 subsets of 4 to which we assign a single focal length value (in pixels) respectively equal to $f_{\min} = 800$, $f_{\max} = 2200$ (see figure 1).

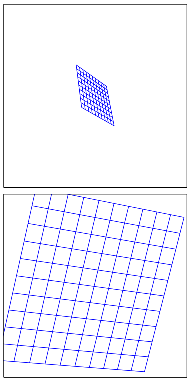

Figure 1: Samples of images associated with $f_{\min}$ and $f_{\max}$.

Regarding the external parameters, for each camera we randomly choose its position in a sphere whose the centre is the centre of the grid and radius is equal to $50cm$. The fixated point by the cameras is roughly the centre of the grid. A random rotation around the optical axis is eventually added with mean 0 and standard deviation 2π. We also restrict the orientation between the 3D plane and the optical axis to the range $[30°, 70°]$ in order to put the simulation in the best situations for the Sturm & Maybank's algorithm[5]. The comparisons focus on error computations at different levels of Gaussian noise, with mean 0 and standard deviation σ varying from 0.5 to 2 pixels, with a step of 0.5. This noise is added to the projected points in the image. Each world to image homography matrix is estimated from the sole visible points in the view using the normalized Direct Linear Transformation (DLT) algorithm [1]. We compute absolute errors (in pixels) for (u_0, v_0), relative errors for κ and mean relative errors $f_{\min}$, $f_{\max}$. Each result, presented in table 1, is the mean of 2000 independent trials. It is worth mentioning that the absolute pixel error function on u_0, v_0 roughly behaves like 7.5σ with the CL-based method and like 10σ with Sturm & Maybank's one. The estimations about κ, $f_{\min}$, $f_{\max}$ are better with our method even if the error on κ is always very low (1.5% at the worst). The fact that relative errors on $f_{\min}$ are higher than errors on $f_{\max}$ can be likely explained by the way we introduce the noise. Indeed, relative distances between points are smaller in images $\mathcal{I}_{\min}$ associated with $f_{\min}$ than in images $\mathcal{I}_{\max}$ associated with $f_{\max}$ (see figure 1). Consequently, for the same amount of noise, the estimation of the homographies is less accurate for $\mathcal{I}_{\min}$ than for $\mathcal{I}_{\max}$.

Real images. We also check our algorithm with 6 real images of a calibration grid taken at a distance of about $40cm$ from different positions. Images are taken by a *Nikon Coolpix® 800* in a 1600×1200 *jpeg* format with the maximal setting of zoom. We know that the bigger zoom is used, the less distortion is present in the images. The camera operate small variations of the focal length during the auto-focusing. In each image, the corresponding centre line is represented by a solid line. The principal point is drawn with a cross. The recovered internal parameters are $u_0 = 810.14$, $v_0 = 558.32$, $\kappa = 1.0026$, $f^{(1)} = 3472.95$, $f^{(2)} = 3507.95$, $f^{(3)} = 3656.54$, $f^{(4)} = 3479.40$, $f^{(5)} = 3402.76$, $f^{(6)} = 3465.67$. The error vector $\frac{\bar{\Phi}\mathbf{x}}{x_4}$, *i.e.* whose elements are orthogonal distances from the estimated principal point to the centre line in each image, has mean 5.5 pixels with standard deviation 5.8. It is

[4] In the following section, notation-wise $x = 200 \pm 10\%$ (d-distributed) means x has a mean value 200 with standard deviation 20 of a d distribution. If d is omitted, the distribution is assumed to be uniform.

[5] Indeed, in their own simulations, they noticed that focal length estimates are slightly better within this range.

worth mentioning that, in image 3, the distance is 15.5 pixels. It confirms the difficulty we had when extracting the grid points in this image ; this is probably due to a “bad” focusing during the acquisition.

Error	abs.		rel. (%)		
	u_0	v_0	κ	$f_{\min}$	$f_{\max}$
$\sigma = 0.5$ (LC)	3.59	3.64	0.25	1.27	0.56
$\sigma = 0.5$ (SM)	4.96	4.93	0.32	1.50	0.75
$\sigma = 1$ (LC)	7.20	7.09	0.49	2.52	1.09
$\sigma = 1$ (SM)	9.87	9.57	0.61	2.97	1.47
$\sigma = 1.5$ (LC)	10.94	10.63	0.73	3.80	1.64
$\sigma = 1.5$ (SM)	14.91	14.30	0.89	4.49	2.20
$\sigma = 2$ (LC)	14.77	14.86	1.00	4.97	2.21
$\sigma = 2$ (SM)	19.76	19.91	1.22	5.92	2.98
$\sigma = 2.5$ (LC)	18.41	18.63	1.26	6.35	2.76
$\sigma = 2.5$ (SM)	24.25	24.95	1.53	7.42	3.68

Table 1: Comparison tests. The CL-based method (CL) *vs* Sturm & Maybank’s (SM) method.

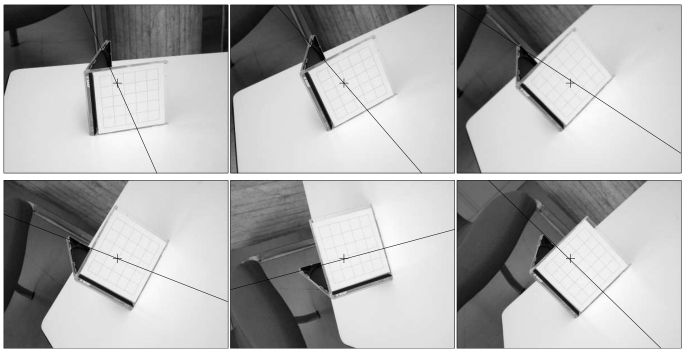

Figure 2: Plane-based calibration from 6 views of a planar grid sticked on a CD jewel case. Centre lines are displayed. The principal point (cross) is the point such that the sum of squared orthogonal distances from it to each center line is minimum.

5 Conclusion

In this paper, we propose a new constraint for plane-based calibration from n images. It particularly applies to case of a varying focal length since this constraint is irrespective of it. The calibration is carried out using a two-step method: at first, the aspect ratio and the principal point are estimated, after which each focal length is directly computed. Our method offers two strong points: [i] the required amount of flops is in $\mathcal{O}(n)$ while it was in $\mathcal{O}(n^3)$ in the existing one, [ii] the results have a better accuracy (at least in simulations). This is probably due to the fact that our cost function has a geometric interpretation involving orthogonal distances and that its minimization is equivalent to a MLE estimation. Regarding the future prospects, we have to compare our experimental results with those of any maximum-likelihood algorithm, in order to get a baseline performance. Our approach shows some obvious limits, notably because the principal point is supposed to be constant. The geometric properties of our linear system enables us to envisage a Kalman-like recursive estimation (with a cost function integrating the evolution of principal point coordinates and focal lengths) taking into account each additional view as soon as it is available.

References

[1] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, UK, 2000.

[2] D. Liebowitz, A. Criminisi and A. Zisserman.*Creating architectural models from images*. In Proc. EuroGraphics, volume 18, pages 39-50, September 1999.

[3] D. Liebowitz and A. Zisserman.*Combining scene and auto-calibration constraints.* In Proc. 7th International Conference on Computer Vision (ICCV'99), Kerkyra, Greece, September 1999.

[4] E. Malis and R. Cipolla. *Self-calibration of zooming cameras observing an unknown planar structure.* In Proc. 15th International Conference on Pattern Recognition (ICPR'00), volume 1, pages 85-88, Barcelona, Spain, September 2000.

[5] C. Matsunaga and K Kanatani. *Calibration of a moving camera using a planar pattern: optimal computation, reliability evaluation, and stabilization by model selection.* In Proc. 6th European Conference on Computer Vision (ECCV'00), Dublin, Ireland, July 2000.

[6] J.V. Poncelet. *Traité des propriétés projectives des figures.* Imprimerie de Mallet-Bachelier, Paris, 1862.

[7] P.F. Sturm and S.J. Maybank. *On plane-based camera calibration: a general algorithm, singularities, applications.* In Proc. Computer Vision and Pattern Recognition Conference (CVPR'99), pages 432–437, Fort Collins, Colorado, USA, June 1999.

[8] B. Triggs. *Autocalibrationfrom planar scenes*. In Proc. 5th European Conference on Computer Vision (ECCV'98), pages 89–105, Freiburg, Germany, June 1998.

[9] B. Triggs. *Géométrie d'images multiples*[6]. Thèse de l'Institut National Polytechnique de Grenoble, November 1999.

[10] M. Wilczkowiak, E. Boyer and P. Sturm. *Camera calibration and 3D reconstruction from single images using parallelepipeds.* In Proc. 12th International Conference on Computer Vision (ICCV'2001), Vancouver, Canada, July 2001.

[11] Z. Zhang. *Parameter estimation techniques: a tutorial with application to conic fitting*. Reseach report INRIA N°2676, October 1995.

[12] Z. Zhang. *A flexible new technique for camera calibration.* IEEE Transactions of Pattern Analysis and Machine Intelligence (PAMI), volume 22, number 11, pages 1330-1334, November 2000.

[6] mainly written in English.