Camera calibration without feature extraction

Luc Robert

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE

Camera Calibration without Feature Extraction

Luc Robert

N° 2204

Fevrier 1994

PROGRAMME 4

Robotique,

image

et vision

ISSN 0249-6399

INRIA
SOPHIA ANTIPOLIS

Camera Calibration without Feature Extraction

Luc Robert*

Programme 4 — Robotique, image et vision
Projet Robotvis

Rapport de recherche n° 2204 — Fevrier 1994 — 23 pages

Abstract: This paper presents an original approach to the problem of camera calibration using a calibration pattern. It consists of directly searching for the camera parameters that best project three-dimensional points of a calibration pattern onto intensity edges in an image of this pattern, *without explicitly extracting the edges*. Based on a characterization of image edges as maxima of the intensity gradient or zero-crossings of the Laplacian, we express the whole calibration process as a one-stage optimization problem. A classical iterative optimization technique is used in order to solve it.

Contrary to classical calibration techniques which involve two consecutive stages (extraction of image features and computation of the camera parameters), our approach does not require any customized feature extraction code. As a consequence, it can be directly used with any calibration pattern that produces image edges, and it is also more robust.

First, we describe the details of the approach. Then, we show some experiments in which two implementations of our approach and two classical two-stage approaches are compared. Tests on real and synthetic data allow us characterizing our approach in terms of convergence, sensitivity to the initial conditions, reliability, and accuracy.

Key-words: Camera, Calibration

(Résumé : tsvp)

This work was partially supported by the EEC under Esprit project. Luc Robert was supported by an INRIA fellowship.

*Luc.Robert@sophia.inria.fr

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Calibration d'une caméra sans extraction de points d'intérêt.

Résumé : Nous penchant sur le problème de la calibration d'une caméra à partir d'une mire, nous présentons dans cet article une approche originale qui consiste à rechercher directement les paramètres de la caméra qui projettent les points du modèle tridimensionnel sur des contours de l'image, *sans extraire ces contours explicitement*. Nous fondant sur une caractérisation des contours comme maxima du gradient de l'intensité ou zéros du Laplacien, nous ramenons le problème de calibration à un simple problème d'optimisation, résolu au moyen d'une méthode d'optimisation itérative classique.

Contrairement aux techniques de calibration classiques qui comportent deux étapes consécutives (extraction des primitives de calibration d'une image de la mire, puis calcul des paramètres de la caméra à partir de ces primitives), notre approche ne nécessite pas de code spécifique destiné à extraire les primitives de calibration. On peut donc l'utiliser directement avec toute mire de calibration (la seule condition étant que la mire produise des contours dans l'image), et elle est aussi plus robuste.

Dans une première partie, nous décrivons les détails de notre méthode. Ensuite, nous présentons des expériences menées sur des données réelles et synthétiques, visant à comparer notre approche avec quelques techniques classiques. Nous étudions en particulier les propriétés de convergence, sensibilité aux conditions initiales, fiabilité et précision.

Mots-clé : Calibration, caméra

1 Introduction

Calibrating a camera means determining the geometric properties of the imaging process, i.e., the transformation that maps a three-dimensional point, expressed with respect to a reference frame, onto its two-dimensional image whose coordinates are expressed in pixel units. This problem has been a major issue in photogrammetry and computer vision for years. The main reason for such an interest is that the knowledge of the imaging parameters allows relating the image measurements to the spatial structure of the observed scene.

Classical calibration techniques, as opposed to recently developed autocalibration techniques [7], proceed by analyzing an image of one or several reference objects whose geometry is accurately known.

These approaches proceed in two steps: First, some features are extracted from the image by means of standard image analysis techniques. These features are generally points or lines, but conics can also be used [12]. Then, they are given as input to an optimization process which searches for the projection parameters that best project the three-dimensional model onto them.

We will not describe in detail the different methods that have been developed. Detailed reviews of the main existing approaches can be found in [13, 14, 16]. We just remark that the approaches can be classified into several categories, with respect to

- the camera model: most existing calibration methods assume that the camera follows the pinhole model. Some of them (mostly in photogrammetry) consider additional parameters that model image distorsions. A good study of the different geometrical distorsion models can be found in [16].

- the optimization process: linear optimization processes are often used in computer vision because they are faster and provide satisfactory results for vision applications. This is possible when the projection equations can be expressed in a linear way, for instance with the pinhole model. When sophisticated camera models are considered, one often needs to use a direct non-linear optimization technique.

It is well known that in a process composed of consecutive stages, numerical errors tend to propagate along the different steps of the process. In the

case of two-stage calibration, if the detected features are inaccurate, then the optimization cannot recover the exact camera parameters. In fact, the calibration process cannot perform better than the less accurate of the two steps, i.e., feature detection and projection parameters recovery.

In order to avoid such a problem, we propose an alternate approach that performs calibration in one step, without introducing the intermediate feature representation. It computes the projection parameters by iteratively minimizing a criterion directly related to the image grey-level intensity.

In the first part of this paper, we describe the camera and grid models that we use. Then, we describe the principles and details of our method. In the last part, we show an experimental study of the behavior of our calibration technique, and compare it to some other classical approaches. Using real and synthetic image sequences, we investigate the following cues: convergence and sensitivity to initial conditions, reliability, and accuracy of the computed parameters. We show that our method compares favorably with some classical two-stage approaches.

1.1 The camera model. Notations

The camera model that we consider is the standard pinhole model. If M has world coordinates $[X_w, Y_w, Z_w]^t$ and projects onto a point m that has pixel coordinates $[u, v]^t$, the operation can be described by the linear equation:

$$\begin{bmatrix} Su \\ Sv \\ S \end{bmatrix} = \tilde{\mathbf{P}} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where S is a scale factor, and $\tilde{\mathbf{P}}$ is the 3×4 projection matrix of the camera.

The projection matrix can be decomposed into the product

$$\tilde{\mathbf{P}} = \tilde{\mathbf{P}}^{\text{cam}}\, \mathbf{K} \tag{2}$$

where:

- $\mathbf{K}$ represents the mapping from world coordinates to camera coordinates. Thus, it accounts for the six extrinsic parameters of the camera (three

for the rotation $\mathbf{R}$, three for the translation $\mathbf{t}$).

$$\mathbf{K} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{O_3}^t & 1 \end{bmatrix} \tag{3}$$

- $\tilde{\mathbf{P}}^{\text{cam}}$ is the matrix of intrinsic parameters, of the following form [6]:

$$\tilde{\mathbf{P}}^{\text{cam}} = \begin{bmatrix} \alpha_u & -\alpha_u \cot\theta & u_0 & 0 \\ 0 & \alpha_v \sin\theta & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{4}$$

  It holds for the five intrinsic parameters of the camera:

  - α_u, α_v are the scale factors along the axes of pixel coordinates,
  - u_0, v_0 are the pixel coordinates of the principal point (orthogonal projection of the optical center on the image plane),
  - θ is the angle of the two axes of pixel coordinates.

The parameter θ allows the representation of any projective transformation from the space to the plane, since it brings the total number of parameters to 11, i.e., the number of degrees of freedom of an homogeneous 3×4 matrix.

1.2 Calibration object

The geometry of the objects that are used for calibration needs to be known very accurately. For this reason, either the calibration objects are specially manufactured in order to conform as accurately as possible to a three-dimensional model, or the model is obtained a-posteriori by measuring the calibration object with some specific devices such as theodolites. The first solution leads to much simpler experiments, but is less accurate. For reasons of time and simplicity of experimentation, it is usually preferred in computer vision. Because they use optical devices of much higher precision, photogrammeters usually prefer the latter solution.

As an example, the calibration object designed and used at INRIA is shown in Figure 1, These are the grid and the reference frame we used for the experiments presented in this article.

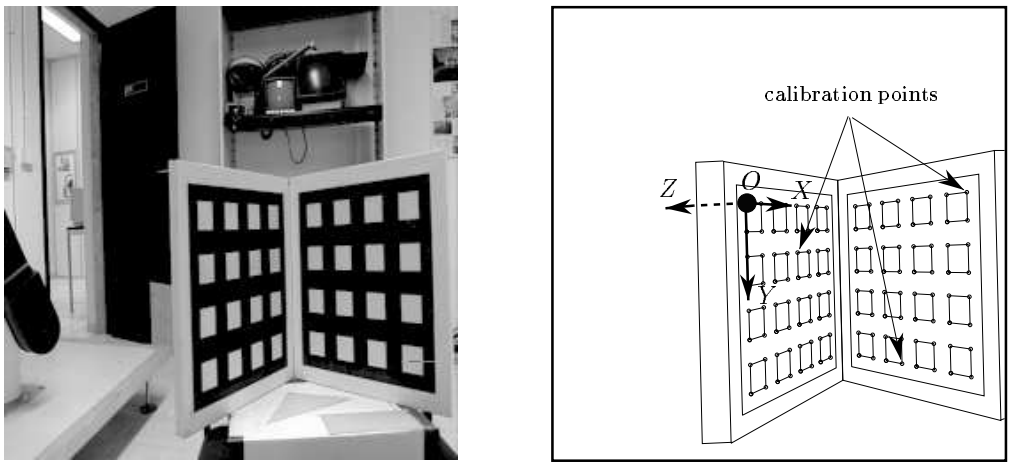

Figure 1: The INRIA calibration grid and the associated geometric model. The model points are the corners of the small square regions. Their coordinates are expressed with respect to the frame represented in the figure.

2 Calibrating from images

Two-stage calibration approaches proceed in the following manner: First, a set A_2 of calibration features are extracted from the images. Then, an optimization process finds the camera parameters that best project the set A_3 of three-dimensional model features onto the extracted image features, by minimizing a distance $d(\tilde{\mathbf{P}}(A_3), A_2)$ between these two sets of features. This distance can be considered as a measure of the edge-ness of the projected features.

The basic idea of our approach is to measure the edge-ness of projected model points directly in the image, without extracting calibration features A_2. For any given set of camera parameters, we can compute an energy value from the image characteristics at the projected model features. Camera calibration consists of minimizing this energy iteratively, in an image-driven process somewhat analogous to a snake [10, 1]. The process starts from an initial value of the projection parameters which is either computed from six reference points (sufficient for deriving the 11 camera parameters) entered manually, or by any existing calibration technique.

2.1 Edge features

As in most of the existing edge detectors, we characterize edges either as zero-crossings of the intensity Laplacian ΔI [11], or as local maxima of the module of the intensity gradient $\|\nabla I\|$ in the direction of the gradient [2, 4]. It has been proved that these models are accurate along regular edges, but not at image corners or junctions [5]. Therefore, the model features we choose are points A_{3_i} lying on three-dimensional edges, as far as possible from any corner or junction. We end up with the two following criteria to be minimized:

1. *Edges are defined as maxima of the intensity gradient:* we maximize

$$\mathcal{C}_{\nabla I}(\tilde{\mathbf{P}}) = \sum_{i \in I} \|\nabla I(\tilde{\mathbf{P}}(A_{3_i}))\|^2$$

2. *Edges are defined as zeros of the Laplacian of intensity:* we minimize

$$\mathcal{C}_{\Delta I}(\tilde{\mathbf{P}}) = \sum_{i \in I} |\Delta I(\tilde{\mathbf{P}}(A_{3_i}))|^2$$

An advantage of this approach is that it can be applied directly to any calibration object that produces edges, while two-stage approaches generally require customized feature extraction code for each target pattern. For instance, the same process can be applied to the two different calibration patterns shown in Figure 2, only the coordinates of the three-dimensional model points differ.

We remark that a point constrained to lie on an edge still still has one degree of freedom (it is allowed to "slide" along the edge), so at least 11 points are required in order to constrain the 11 parameters of the projection matrix.

2.2 Parameterization of $\tilde{\mathbf{P}}$

The projection matrix can be represented by the set of 11 intrinsic and extrinsic parameters:

- 5 numbers for $\alpha_u, \alpha_v, u_0, v_0, \theta$
- 3 numbers t_X, t_Y, t_Z for the translation $\mathbf{t}$

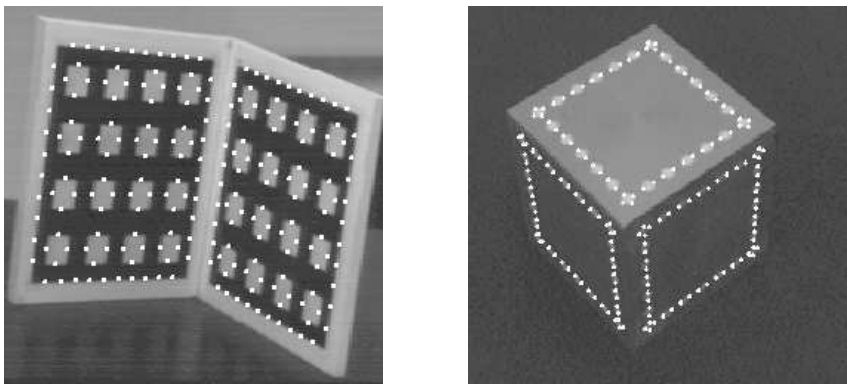

Figure 2: Calibration objects and points of the model A_3 for the INRIA grid (left) and the CMU cube (right).

- 3 numbers n_X, n_Y, n_Z for the rotation $\mathbf{R}$ ($\mathbf{n}$ defines the direction of the axis of rotation, and its module is the angle of the rotation).

Another representation consists of setting one of the coefficients of $\tilde{\mathbf{P}}$ to 1, after checking that this coefficient is non zero around the solution. The matrix is then represented by its other 11 coefficients.

The choice of this coefficient is guided by the following remark: Eq. (2), (3) and (4) show that p_{34} (the bottom-right coefficient of $\tilde{\mathbf{P}}$) is equal to the third component of the translation vector $\mathbf{t}$. Given the choice of the reference frame (cf. Figure 1) and the position of the camera that observes the grid, p_{34} cannot be zero if $\tilde{\mathbf{P}}$ is close to the actual projection matrix. Thus, we set $p_{34} = 1$ and define $\tilde{\mathbf{P}}$ with respect to this coefficient.

2.3 Implementation details

The minimization procedure we use is a Newton-type iterative method. In practice, we use a gradient descent routine of the *NAG* [9] library. The estimate of the gradient of the criterion is obtained by finite difference approximations. We evaluate subpixel intensity derivatives at a given point by means of bilinear interpolation of the values at the four adjacent pixels.

An example of execution of the algorithm is shown in Figure 3. On average, convergence requires 20 iterations (numerical estimations of the gradient) with

a maximum of 400 elementary descent steps at each iteration. Computation time remains quite reasonable: 15 seconds on a *Sun Sparc 2* workstation for a model containing 208 points.

3 Experimental results

In this section, we show the results of experiments conducted in order to compare two implementations of our method (based on image gradient or Laplacian) with two classical two-stage methods, one using direct linear optimization, the other using non-linear iterative minimization. We study the influence of the initial estimate of the projection matrix on the resulting parameters. Then we compare the reliability of the four methods in the estimation of intrinsic parameters, and relate it to the recovery of extrinsic parameters.

3.1 Influence of the initial conditions

Two essential points in the study of an iterative process are the influence of the initial conditions and the convergence properties. In order to study these two points, we use a real image and generate a set of initial data that are close to the solution. By measuring the variation of the results produced by three iterative methods, we compare their convergence properties. It is important to notice that the exact solution of the calibration process is unknown, but this is not really a problem since we are only interested in the consistency of the results.

The initial projection matrices are determined in the following manner: a set of calibration points are extracted from the image; using these points with the Faugeras-Toscani method, we compute a good estimate of the projection matrix. Since we want to perturb this matrix, we add some noise to the points, and use the noisy data as input to the Faugeras-Toscani method. The computed projection matrices are used as initial values for the iterative algorithms. Average and standard deviation of the parameters of the initial matrices are represented in the last two columns of the top two tables in Table 1.

We compare the three following iterative methods:

- maximization of the image gradient (denoted by (GR)),

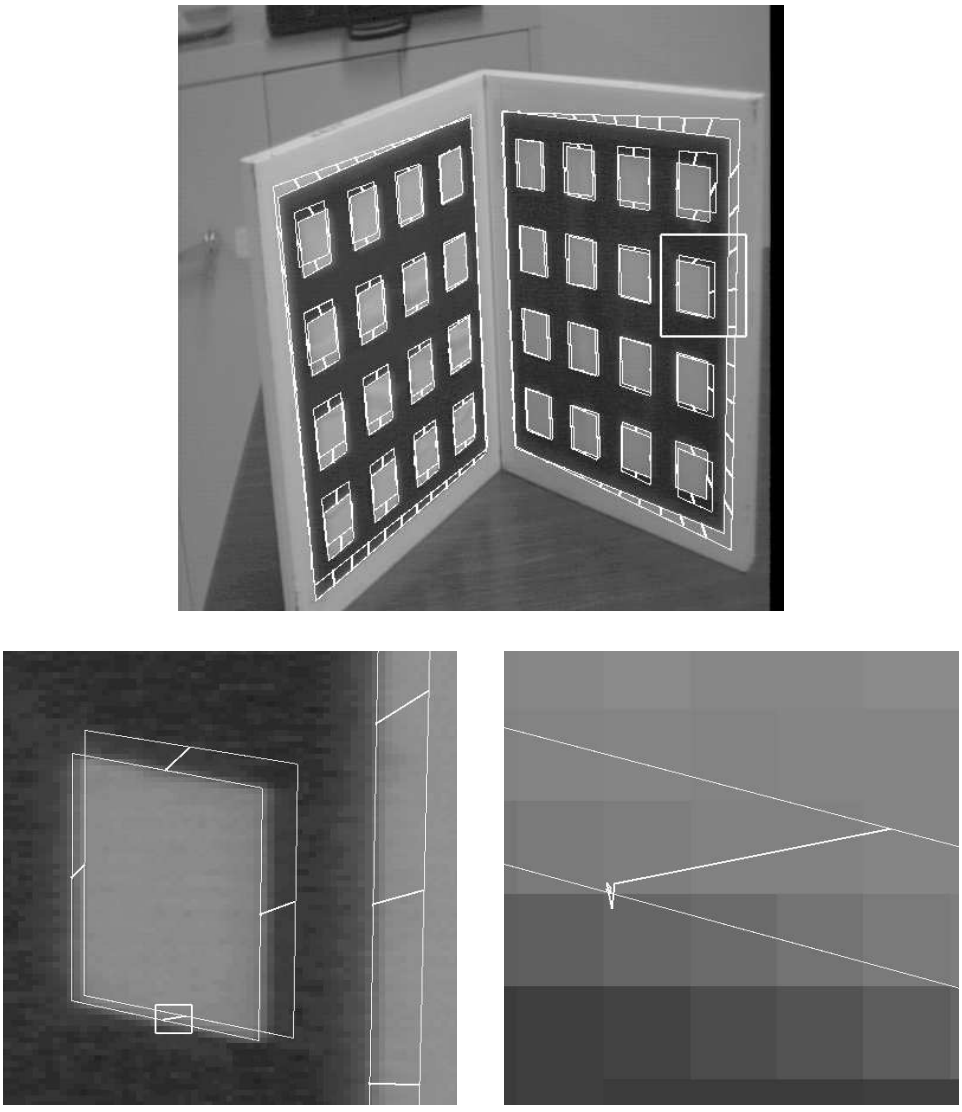

Figure 3: Evolution of the iterative calibration process: the top image represents projections of the 3D model grid onto the calibration image, with both initial and final estimates of the projection matrix. The trajectories of the model points of A_3 are also represented on the image. Both bottom images show enlargements of sub-regions of the image, outlined by white rectangles in top and bottom-left images.

- minimization of the image Laplacian (denoted by (LAP)),
- minimization of the distance between the projected points and the non-perturbated calibration points [15] (denoted by (DIST)).

In the following, we will refer to the Faugeras-Toscani method as (FT).

The average and standard deviation of the parameters computed for 50 different initial matrices are given in Table 1. For the first table, calibration points are slightly perturbated (a gaussian noise with a standard deviation of 0.2 pixel is added to the points). In the second table, the amplitude of the added noise is 2 pixels. The columns of the first two tables represent the averages and standard deviations on the parameters computed with the three iterative methods. By c we denote the cosine of the angle θ between the pixel axes. The third table shows the results of the Faugeras-Toscani method when the points are not perturbated.

We observe the following properties :

- Overall, the values of standard deviations are very small with respect to the camera parameters, and with respect to the standard deviations on the initial parameters. In other words, *the influence of the initial value is very limited.* This is confirmed by the fact that average values and standard deviations are comparable from one table to the other, even though the level of noise is not the same. This property is very important, because without it, the method could not be used reliably.
- The standard deviations for the (LAP) method are much smaller than for the (GR) and (DIST) methods. In other words, the optimum of the criterion involving the Laplacian is sharper. This seems advantageous, since the solution is computed with better accuracy, and the process converges faster. However, another consequence is that the algorithm is more likely to end in a local minimum, so *the (LAP) method has to be initialized closer to the solution than (GR) and (DIST).* We actually noticed a few cases where the (LAP) process converged to a local minimum, when a 2-pixel noise was added to the starting points. These cases were not included in the statistics.
- The differences between the average values computed for the three iterative methods and for the (FT) method are small. But comparison with

	initial noise : 0.2 pixel							
	gradient (GR)		*Laplacian (LAP)*		*points (DIST)*		*initial*	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
α_u	772.01	1.14	763.94	0.238	774.16	0.922	768.13	2.23
α_v	1175.2	1.65	1163.9	0.345	1178.5	1.35	1169.7	3.25
u_0	262.54	0.609	262.32	0.162	259.11	0.263	257.5	0.781
v_0	275.36	1.12	269.63	0.393	278.14	0.5	276.06	1.76
c	-3.8e-07	1.3e-07	-4.9e-07	8.8e-08	-2.1e-07	5.1e-08	-2.8e-07	2.8e-07
r_X	-1.5995	0.005	-1.5775	0.00164	-1.6206	0.00198	-1.6135	0.00688
r_Y	4.9287	0.00241	4.9383	0.000737	4.9196	0.000909	4.9228	0.00325
r_Z	1.166	0.00357	1.1578	0.000771	1.1842	0.00106	1.1839	0.00374
t_X	-516.07	1.01	-509.24	0.155	-516.76	0.76	-511.92	1.91
t_Y	-201.82	0.48	-199.06	0.0704	-203.36	0.342	-201	0.892
t_Z	-655.93	0.755	-650.21	0.19	-658.55	0.739	-653.84	1.65

	initial noise : 2 pixels							
	gradient (GR)		*Laplacian (LAP)*		*points (DIST)*		*initial*	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
α_u	771.85	1	763.97	0.154	774.53	1.13	648.34	12.4
α_v	1174.9	1.36	1163.9	0.305	1179.1	1.61	994.78	19
u_0	262.72	0.594	262.28	0.159	259.15	0.283	250.91	5.94
v_0	274.84	1.08	269.67	0.41	278.16	0.59	249.96	18.3
c	-3.4e-07	1.3e-07	-4.9e-07	9.6e-08	-1.9e-07	6.1e-08	-1.5e-06	3.7e-06
r_X	-1.5971	0.00467	-1.5777	0.00174	-1.6208	0.00223	-1.5037	0.086
r_Y	4.9299	0.00218	4.9383	0.0008	4.9195	0.00101	4.9757	0.0378
r_Z	1.1644	0.00313	1.1579	0.000757	1.1843	0.00105	1.1502	0.0384
t_X	-515.98	0.947	-509.25	0.132	-517.03	0.952	-408.32	10.7
t_Y	-201.72	0.396	-199.06	0.0558	-203.51	0.436	-154.62	5.87
t_Z	-655.79	0.563	-650.21	0.147	-658.88	0.846	-567.4	10.3

Faugeras-Toscani (FT)			
intrinsic		*extrinsic*	
α_u	769.36	r_X	-1.6156
α_v	1171.5	r_Y	4.9217
u_0	257.72	r_Z	1.1842
v_0	276.66	t_X	513.03
c	-2.3e-07	t_Y	-201.50
		t_Z	-654.71

Table 1: Experiment (a) : influence of the initial conditions.

the standard deviations tends to prove that *these differences are systematic*. Further experiments are necessary in order to find out which methods are biased, and what could be the reason of the bias.

- We notice, by comparing the last two columns of the first two tables, that the parameters yielded by the (FT) method vary systematically with the level of noise added to the points. This tends to show the (FT) method introduces a bias in the estimation of the camera parameters, that becomes detectable if the level of noise on the calibration features increases.

3.2 Study of the intrinsic parameters

The intrinsic parameters should, by definition, remain unchanged if the position of the camera with respect to the grid changes. Thus, in a first step we study the variations of the intrinsic parameters that can be extracted from a sequence of images acquired while the camera is moving with respect to the grid. The average values and standard deviations computed for a sequence of 30 grid images (cf. Figure 4) are given in Table 2.

Based on these results, we make the following remarks:

- While yielding comparable values for u_0, v_0, the different methods produce significantly different estimates of α_u, α_v. However, the ration α_u/α_v is almost constant. In other words, all the methods yield consistent results as far as the shape of the pixels is concerned, but not for the size of the pixels.

- The standard deviations computed with the four different programs have the same orders of magnitude, but the methods that use calibration points ((DIST) and (FT)) yield somewhat less accurate results. The Laplacian gives the most accurate results. The fact that v_0 is determined with a much better accuracy than u_0, already pointed out in [15], has never been really explained.

- The relative differences between the values of intrinsic parameters yielded by the image-based ((GR) or (LAP)) and point-based ((DIST) (or (FT)) methods are much bigger than in the data of Table 1. We believe that this

	real sequence					
	α_u		α_v		α_u/α_v	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
gradient (GR)	937.52	14.5	1388.8	20.8	0.6750	0.0010
Laplacian (LAP)	920.56	6.59	1364.6	9.16	0.6746	0.0008
points (DIST)	849.35	18.6	1254.1	26.8	0.6772	0.0013
Faugeras-Toscani (FT)	831.23	23.1	1229.4	33.4	0.6761	0.0010

	real sequence					
	u_0		v_0		c	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
gradient (GR)	272.51	11	205.93	13.1	2.1e-06	1.1e-06
Laplacian (LAP)	269.28	13.8	226.78	4.26	2.1e-06	5.4e-07
points (DIST)	266.23	20.9	220.45	10.3	9.1e-07	1.0e-06
Faugeras-Toscani (FT)	258.91	22.3	212.01	11	1.5e-06	1.0e-06

Table 2: Experiment (b) : study of intrinsic parameters on a real sequence.

is related to the fact that the grid lies at a longer range from the camera, and the experiment that we describe in the next paragraph illustrates this fact.

It is impossible to know the exact values of the parameters of a real camera. In order to compare the accuracy of the different methods, we generate a sequence of synthetic images, that simulates what a camera moving around the calibration grid would produce. The camera is represented by the matrix of intrinsic parameters $\tilde{\mathbf{P}}^{\text{cam}}$ (Eq. (2) and (4)) which remains by definition unchanged for all the sequence. The consecutive positions of the camera are defined by a sequence of displacement matrices $\mathbf{K}_t$. By means of a ray-tracing program, we projected a synthetic model of the INRIA grid on the virtual cameras of projection matrices $\tilde{\mathbf{P}}_{\text{cam}}\mathbf{K}_t$. Each image of the sequence is then corrupted by a gaussian noise. The results obtained for the two image sequences are shown in Table 3.

- Globally, the (GR) and (LAP) methods provide quite accurate estimates of the intrinsic parameters. The fact of adding noise to the images has a very small affect on the results of (GR). The (LAP) method is seemingly

less robust, even though the error on the estimated parameters is close to 1% for the sequence of noisy images.

- The (DIST) and (FT) methods give less accurate results.
- Once again, the ratio α_u/α_v does not depend on the method that is used.

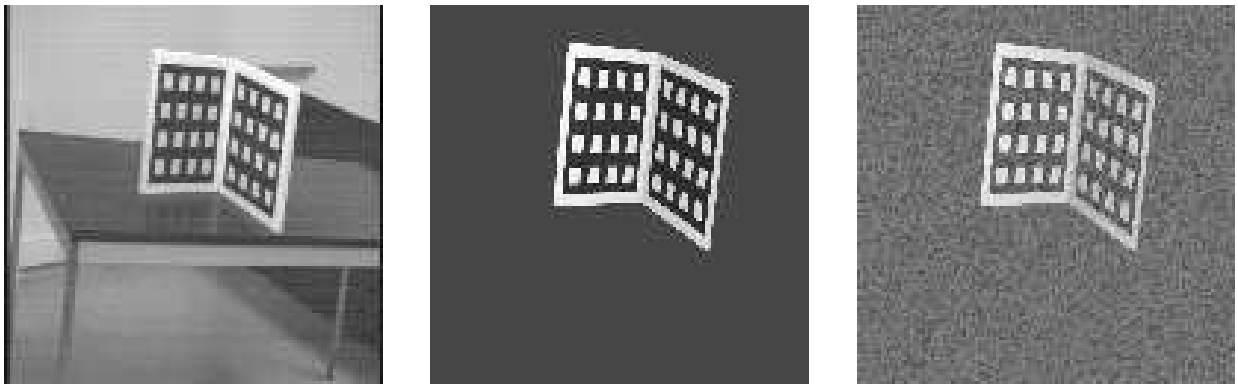

Figure 4: Examples of images from real, synthetic and noisy sequences.

3.3 Relation with the extrinsic parameters

Assuming that pixel axes are orthogonal ($\theta = \frac{\pi}{2}$), Eq. (1) and (4) yield

$$\begin{cases} \alpha_u \dfrac{X_c}{Z_c} &= u - u_0 \\ \alpha_v \dfrac{Y_c}{Z_c} &= v - v_0 \end{cases}$$

When M has depth Z_c^0, its projection on the retina is characterized by the ratios $\frac{\alpha_u}{Z_c^0}, \frac{\alpha_v}{Z_c^0}$. As a consequence, if we consider that the points of the grid lie at approximately the same depth (which is even more valid when the grid is far from the cameras) the estimation of the pixel sizes (α_u, α_v) must be directly correlated to the estimation of the distance between the grid and the camera. In order to investigate this, we acquire a sequence of 15 images in which the grid is a various distances from the camera. For each image, we measure the

	synthetic sequence without noise					
	α_u		α_v		α_u/α_v	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
nominal	940.00	0.00	1400.00	0.00	0.6714	0.00
gradient (GR)	940.45	10.2	1400.6	15.4	0.6715	0.0004
Laplacian (LAP)	939.57	3.3	1399.5	4.86	0.6714	0.0002
points (DIST)	951.6	10.2	1416	14.5	0.6720	0.0005
Faugeras-Toscani (FT)	932.87	8.09	1388.7	11.7	0.6718	0.0004

	synthetic sequence without noise					
	u_0		v_0		c	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
nominal	270.00	0.00	205.00	0.00	0.00	0.00
gradient (GR)	270.47	3.94	205.32	8.99	1.5e-07	6.62e-07
Laplacian (LAP)	269.94	2.08	205.08	3.23	-1.3e-08	2.1e-07
points (DIST)	270.31	5.78	209.33	9.4	-4.7e-07	5.2e-07
Faugeras-Toscani (FT)	270.37	5.27	207.28	8.39	-2.9e-07	4.6e-07

	synthetic sequence with noise					
	α_u		α_v		α_u/α_v	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
nominal	940.00	0.00	1400.00	0.00	0.6714	0.00
gradient (GR)	941.73	11.8	1402.4	17.6	0.6716	0.0005
Laplacian (LAP)	934.4	5	1392	7.13	0.6713	0.0003
points (DIST)	950.87	15.1	1415.4	22.4	0.6718	0.0005
Faugeras-Toscani (FT)	928.3	7.9	1382.3	11.8	0.6715	0.0004

	synthetic sequence with noise					
	u_0		v_0		c	
	average	*std. dev.*	*average*	*std. dev.*	*average*	*std. dev.*
nominal	270.00	0.00	205.00	0.00	0.00	0.00
gradient (GR)	269.96	5.32	204.04	9.92	1.2e-07	7.5e-07
Laplacian (LAP)	270.39	2.98	203.17	5.4	8.6e-08	4.4e-07
points (DIST)	268.9	7.55	204.51	10.8	-5.3e-07	5.3e-07
Faugeras-Toscani (FT)	270.12	5.89	203.07	8.72	-4.3e-07	5. 1e-07

Table 3: Experiment (c) : intrinsic parameters for the sequence of synthetic images, with and without noise.

distance between the optical center of the camera and the origin of the frame on the grid. Figure 5 shows three images of the grid sequence, including the closest and the furthest ones.

We denote by $d_{\mathrm{m,c}}$ the distance from the origin of the grid to the optical center, and by d_{man} the value of $d_{\mathrm{m,c}}$ measured manually. Figure 6 represents the values of $d_{\mathrm{m,c}}$ computed with different calibration methods, as functions of d_{man}. Figure 7 shows the values of α_u (thin lines) and of $\frac{\alpha_u}{d_{\mathrm{m,c}}} d_{\mathrm{man}}$ (thick lines) as functions of d_{man}.

- At short range, all the methods estimate $d_{\mathrm{m,c}}$ correctly (up to 1.30m, the highest error is 1%). The values for α_u are also close.
- If the distance between the grid and the camera increases, the values of $d_{\mathrm{m,c}}$ given by methods (DIST) and (FT) start differing noticeably and systematically from the manually measured values. The (GR) and (LAP) methods give results that remain much closer to these values. There is also a noticeable gap between the values of α_u computed with the two-stage and the one-stage calibration methods. This gap increases with the distance between the grid and the camera.
- Since the origin of the grid is relatively close from the optical axis (the maximum distance is of the order of 20cm), $d_{\mathrm{m,c}}$ is a good estimate of the depth of the grid (the relative error is of the order of 3% if the origin of the grid is 90cm away from the camera, less than 1% if it is 1.50m away). d_{man} is also very close to the actual value. Figure 7 illustrates the above assertion that the error on $d_{\mathrm{m,c}}$ is compensated by an error on α_u. Indeed, even if the estimate of α_u varies with d_{man} for the four methods (thin lines), the parameter $\frac{\alpha_u}{d_{\mathrm{m,c}}} d_{\mathrm{man}}$ remains almost constant (thick lines), all the methods yielding approximately equal values.

In their study of the evolution of the results of (DIST) when calibration points are perturbated with isotropic noise from their nominal position, Chaumette [3] and Vaillant [15] did not notice any bias on the results. This tends to show that the coupled errors in the estimation of the intrinsic and extrinsic parameters are not directly related to the determination of $\tilde{\mathbf{P}}$ itself, but to the first steps, which introduce a bias in the position of the calibration points. This

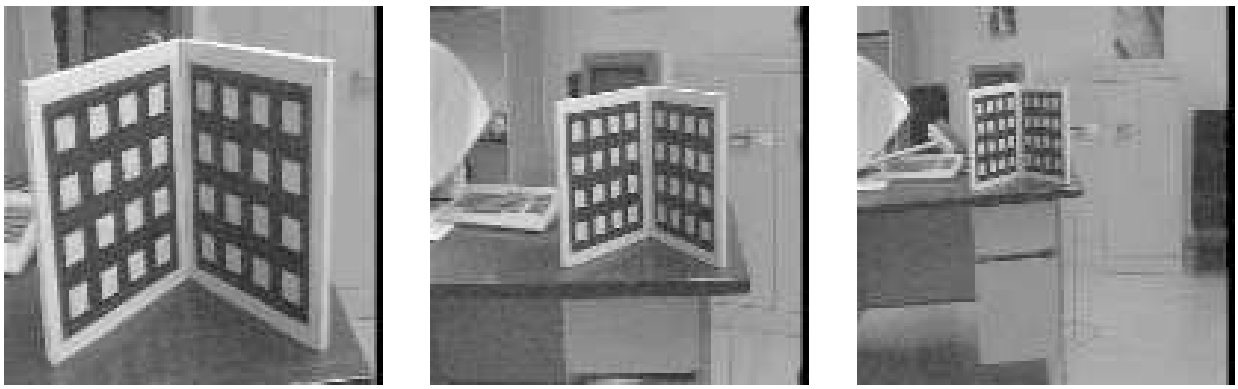

Figure 5: Experiment (d): Three images of the sequence.

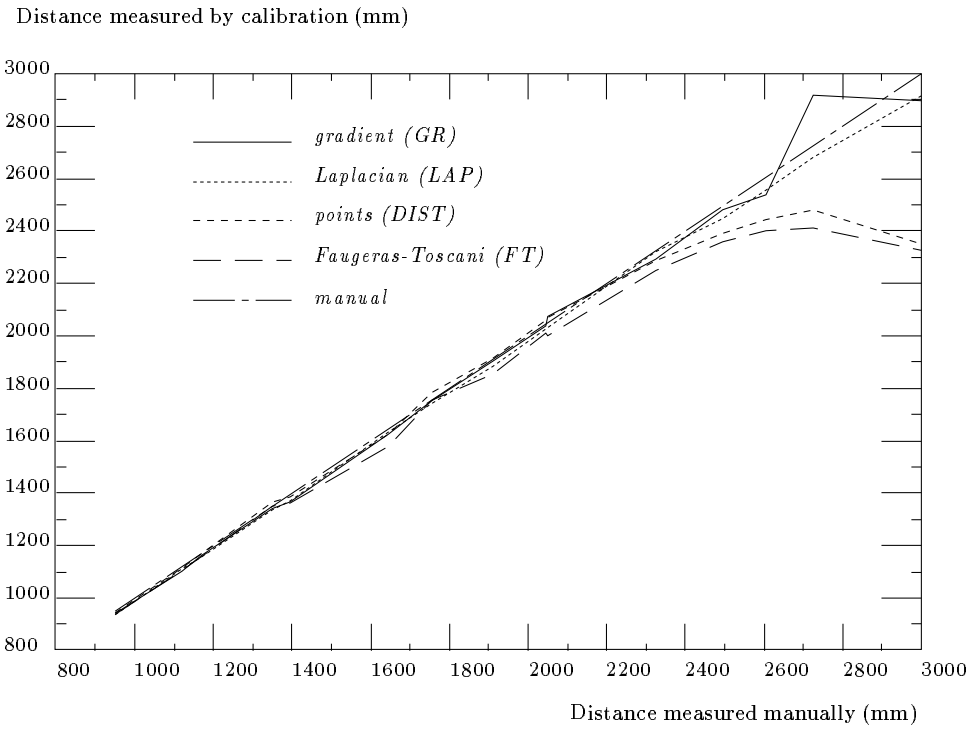

Figure 6: Experiment (d): distances evaluated with different calibration methods.

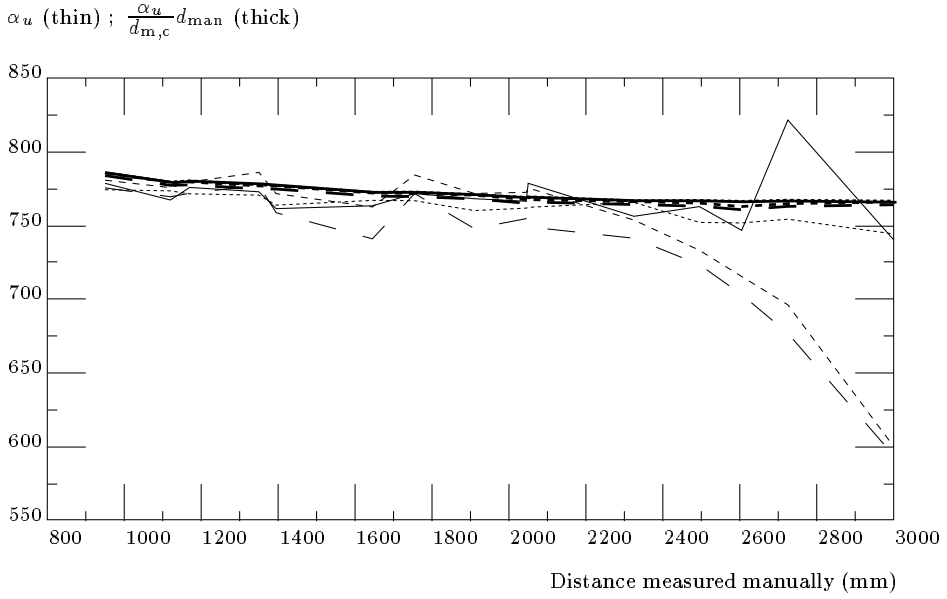

Figure 7: Experiment (d): correlation of α_u and $d_{\mathrm{m,c}}$.

could either be caused by the discretization performed while extracting edges from the image, or by the imperfections in the construction of the physical grid. At any rate, it illustrates that even though the search for the projection parameters is unbiased for an isotropic perturbation of the input features, it can provide biased results because the output of the feature detector is itself biased. Our one-stage approach does not suffer from this problem.

4 Conclusion

In this article, we present an method for camera calibration which is different from the classical ones, in that it does not extract any image feature but uses the grey-levels image itself.

Contrary to classical calibration techniques, our approach can be directly used with various types of calibration objects. It can be initialized by hand, or by a process that recognizes a small number of model points in the image. Therefore, it is easier to use than most classical calibration techniques, which rely on a complicated feature-extraction process that depends on the shape of the calibration object that is used and on the kind of features that are extracted.

From a numerical standpoint, we experiment with two different formulations of the approach, one based on the maximization of the image intensity gradient (GR), the other on the minimization of the image Laplacian (LAP). We compare them to two feature-based methods, one that minimizes a non linear criterion [15] (DIST), the other that minimizes a linear criterion [8] (FT). The results are satisfactory, and the main conclusions are the following:

- In a first experiment (a), we point out the good convergence and stability properties of the three iterative methods, i.e., (GR), (LAP) and (DIST). The (LAP) method is somewhat less stable that the others, and has to be initialized closer to the solution.

- The two methods based on calibration points yield noticeably biased estimates of α_u, α_v, and the translation parameters. The bias increases with the distance between the grid and the camera (experiment (d)). Our one-stage methods do not introduce such a noticeable systematic error

(experiments (a) and (d)), which is due to the feature-extraction process (experiment (d)).

- Minimizing the image Laplacian (LAP) seems more reliable than maximizing the image gradient (GR), but it is slightly biased when the image is corrupted by noise (experiments (c) and (d)).

Though we have only considered the case of pinhole cameras, we believe that the approach can be easily adapted to other camera models that would, for instance, include image distorsion. It is also suited to the case when one or several camera parameters are already known and do not need to be computed (for instance, one can impose the pixel axes to be orthogonal, or can consider that the intrinsic parameters do not change between several experiments). If the intrinsic parameters of the camera are known, the same method can be used to recover the pose of an object. Further work is needed in order to investigate all these issues.

References

[1] B. Bascle and R. Deriche. Stereo Matching, Reconstruction and Refinement of 3D Curves Using Deformable Contours. In *Proc. International Conference on Computer Vision*, pages –, May 1993.

[2] J.F. Canny. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:769–798, November 1986.

[3] F. Chaumette and P. Rives. Modélisation et Calibration d'une caméra. In *AFCET*, pages 527–536, 1989.

[4] R. Deriche. Using Canny's Criteria to Derive an Optimal Edge Detector Recursively Implemented. In *The International Journal of Computer Vision*, volume 2, pages 15–20, April 1987.

[5] R. Deriche and G. Giraudon. A Computational Approach for Corner and Vertex Detection. *The International Journal of Computer Vision*, 10(2):101–124, 1993.

[6] O.D. Faugeras. *Three-Dimensional Computer Vision: a geometric viewpoint.* MIT Press, 1993.

[7] 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 Margherita Ligure, Italy, May 1992. Springer Verlag.

[8] O.D. Faugeras and G. Toscani. The Calibration Problem for Stereo. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 15–20, 1986. Miami Beach, Florida.

[9] The Numerical Algorithms Group. *NAg Fortran Library, Mark 14.* The Numerical Algorithms Group Limited, 1990.

[10] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *The International Journal of Computer Vision*, 1(4):321–331, January 1988.

[11] D. Marr and E. Hildreth. Theory of Edge Detection. *Proceedings of the Royal Society London, B*, B 207:187–217, 1980.

[12] C.A. Rothwell, A. Zisserman, C.I. Marinos, D.A Forsyth, and J.L. Mundy. Relative Motion and Pose from Arbitrary Plane Curves. *Image and Vision Computing*, 10(4):250–262, 1992.

[13] R.Y. Tsai. A Versatile Camera Calibration Technique for High Accuracy 3D Vision Metrology using Off-the-shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*, 3(4):323–344, 1987.

[14] R.Y. Tsai. Synopsis of Recent Progress on Camera Calibration for 3D Machine Vision. In Oussama Khatib, John J. Craig, and Tomás Lozano-Pérez, editors, *The Robotics Review*, pages 147–159. MIT Press, 1989.

[15] R. Vaillant. *Géométrie Différentielle et Vision Par Ordinateur: Détection et Reconstruction des Contours d'Occultation de la Surface d'un Objet Non-Polyédrique*. PhD thesis, Université de Paris-Sud Centre d'Orsay, December 1990.

[16] J. Weng, P. Cohen, and M. Herniou. Camera Calibration with Distorsion Models and Accuracy Evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, 1992.

Unité de recherche INRIA Lorraine, Technôpole de Nancy-Brabois, Campus scientifique,
615 rue de Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, IRISA, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399