From Projective to Euclidean Space under any practical situation, a criticism of self-calibration

Sylvain Bougnoux
INRIA, 2004 route des Lucioles, BP 93
F-06902 Sophia-Antipolis Cedex
Email : Sylvain.Bougnoux@sophia.inria.fr

Keywords

Kruppa equations, Projective versus Euclidean calibration, Projective basis, Focal length, 3D Reconstruction.

Abstract

For many practical applications it is important to relax the self-calibration conditions to allow for changing internal camera parameters (e.g. zooming/focusing ...). Classical techniques failed for such conditions. We present the available constraints that allow us to right a projective calibration to a Euclidean one. Meanwhile, we found that the estimations of the internal parameters were rather inaccurate. We discuss theoretically this difficulty and above all the resulting effect on the 3D reconstruction. In fact, we show that the uncertainty on the focal length estimation leads to a Euclidean calibration up to a quasi anisotropic homothety whereas the error on the principal point can often be interpreted as a translation. Hopefully, the calibration we come up with, is quite acceptable for reconstruction of models.

1 Introduction

Calibration is a crucial stage when trying to get 3D information out of images. The calibration problem has been addressed for many years in Photogrammetry [GB92] and Computer Vision [LF97, MF92]. Most of the proposed approaches consist of two stages: first, the internal parameters are computed off-line using calibration targets; second, using additional information such as control points in the scene, or techniques such as structure from motion, the relative displacements of the cameras are estimated. Self-calibration methods, which allow estimating the internal camera parameters directly from the images, have also been developed [ZF96]. They mainly consist of resolving the well known Kruppa equations under the assumption that the internal parameters are constant. Another approach is to enforce certain movements of the camera, which allow an accurate intrinsic calibration [VF96, Ste95, AZB94]. Several stratified calibration levels have been introduced (projective, affine or Euclidean) [Fau94]: depending on the application, calibration is required only up to a certain level. At least some people [PVG97, HÅ97, Tri97, Har92] tackled the calibration problem without assuming constant internal parameters. Those methods are mainly based on the recovery of the absolute quadric. For the time being they lack good initialization guess and have been proved very tough. We will explain here the purpose of this difficulty.

We have a sequence of images without any particular information of a mostly static scene. We match points along the sequence as described in [BR97] and obtain the epipolar geometries pair-wise. Then, we run bundle adjustment's techniques to calibrate the cameras [Lav96]. Depending on the parameterization we use for our cameras we obtain a projective or Euclidean calibration.

In order to get an initialization for the Euclidean calibration, we use the Kruppa equations assuming constant internal parameters for our cameras. We observed from a practical point of view, as many people did, that self-calibration processes are rather unstable, and even if they have been lead with caution, they might provide really different results. It is really frustrating to get a projective calibration (*i.e.* according to the residual) and no Euclidean one at all (the process doesn't converge).

So what kind of information can we get from the projective space? The geometric parameters reside within it, how can we extract them? In the projective space, we can change from a basis to any other one through a collineation. So following Hartley[Har94], let's see the Euclidean basis as a particular projective basis for which some constraints must be observed. Looking for the Euclidean basis is equivalent to look for a collineation that sends our current projective projection matrices to a set of Euclidean ones.

We present in section 2.1 constraints that must be observed over any Euclidean projective matrix and write them down for the collineation we are looking for. Then, we show in section 3 some experiments in real situations with a set of images where classical methods failed. Finally, we discuss theoretically in section 4 the relationship between the computed internal parameters and the reconstructed scene. In fact, their estimations are fairly unstable, but as many people observed [ZLF96] the scene's structure doesn't depend that much on those parameters and the reconstructed model may always look fine. This is what we demonstrate here.

2 Looking for a Euclidean basis

2.1 The constraints

As we are working in projective space, we represent our projection matrices $\mathbf{P}^i_{proj}$ by 3×4 matrices up to a scale factor. We classically represent any Euclidean projection matrix $\mathbf{P}^i_{euc}$ as the product of $\mathbf{A}$ a 3×3 matrix, holding the internal parameters, $\mathbf{P}_0$ the canonic projection matrix and $\mathbf{T}$ a 4×4 matrix holding the external parameters [Fau93].

$$\mathbf{P}_{euc} = \mathbf{A}\mathbf{P}_0\mathbf{T}$$

$$\mathbf{P}_{euc} = \begin{bmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_3^T & 1 \end{bmatrix}$$

With this formulation, α_u and α_v respectively stand for the scale aspect in the image of the x and y axis, γ is a skew factor, whereas u_0 and v_0 are the image coordinate of the principal point. In almost every practical case we can assume:

$$\begin{aligned} \gamma &= 0 \\ \alpha_u &= \alpha_v \end{aligned} \tag{1}$$

The first approximation has been mostly used for years whereas the second one has become more and more reasonable as the cameras have become more and more accurate[1]. Let's call f α_u and α_v. Those equations have been seen so far as approximations, here we want to see them as constraints over our Euclidean projective matrices.

We have for each 3D point M_j projected onto m^i_j in image i, whatever $\mathbf{Q}$, a nonsingular 4×4 matrix, is:

$$m^i_j = \mathbf{P}^i_{proj}\mathbf{Q}\mathbf{Q}^{-1}M_j \tag{2}$$

[1] If the aspect ratio $\alpha_u/\alpha_v \neq 1$, the following discussion can easily be derived as soon as the aspect ratio is known

Suppose there exists a unique $\mathbf{Q}$ such that for each image i,

$$\mathbf{P}^i_{euc} \simeq \mathbf{P}^i_{proj}\mathbf{Q} = \begin{bmatrix} \tilde{\mathbf{p}}_1^{iT} & \\ \tilde{\mathbf{p}}_2^{iT} & \tilde{\mathbf{t}}_i \\ \tilde{\mathbf{p}}_3^{iT} & \end{bmatrix}$$

Thus assuming we have computed a set of projective projection matrices, we are just looking for $\mathbf{Q}$ a 4×4 matrix, representing the basis transformation, under the previous constraints (1). Whatever $\mathbf{Q}$ is[2], we have a unique *Cholesky decomposition* for any $\mathbf{P}^i_{euc}$. That is to say we have from [Fau93, p53],

$$\gamma_i = 0 \Leftrightarrow (\tilde{\mathbf{p}}^i_1 \wedge \tilde{\mathbf{p}}^i_3).(\tilde{\mathbf{p}}^i_2 \wedge \tilde{\mathbf{p}}^i_3) = 0$$

$$\alpha^i_u = \alpha^i_v \Leftrightarrow \|\tilde{\mathbf{p}}^i_1 \wedge \tilde{\mathbf{p}}^i_3\| = \|\tilde{\mathbf{p}}^i_2 \wedge \tilde{\mathbf{p}}^i_3\|$$

Those constraints are of degree 4, and we give them to a non-linear minimization process, *e.g.* a *Levenberg-Marquardt algorithm.* With such constraints, the internal parameters f_i, u^i_0 and v^i_0 can vary from one camera to another, each camera has 9 unknowns. Now let's see how we can constraint $\mathbf{Q}$ such that for each image i, $\mathbf{P}^i_{euc}$ satisfies our constraints.

2.2 Looking for Q

We assume that we have computed a set of projective projection matrices. Without any particular assumption we can choose $\mathbf{P}^0_{proj} = \begin{bmatrix} \mathbf{I} & \mathbf{0}_3 \end{bmatrix}$ and $\mathbf{P}^0_{euc} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{0}_3 \end{bmatrix}$. So we have:

$$\mathbf{P}^0_{euc} = \mathbf{P}^0_{proj}\mathbf{Q} \Leftrightarrow \begin{bmatrix} \mathbf{A}_0 & \mathbf{0}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0}_3 \end{bmatrix}\mathbf{Q}$$

$$\Leftrightarrow \exists\, (\mathbf{q}, q_{44}) \mid \mathbf{Q} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{0}_3 \\ \mathbf{q}^T & q_{44} \end{bmatrix}$$

Moreover, $\mathbf{Q}$ is defined up to a scale factor and q_{44} appears to be the scale of the scene we are looking at. Thus we can choose:

$$\mathbf{Q} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ q_1 & q_2 & q_3 & 1 \end{bmatrix}$$

Under this parameterization, $\mathbf{Q}$ is clearly nonsingular, and there remains only 6 unknown parameters to define it: f, u_0 and v_0 (which are in fact the unknown internal parameters of the first camera), q_1, q_2 and q_3. As we have 6 unknowns, 2 constraints per image excluding the first one, we need at least 4 images $(1 + 6/2)$ to compute such a $\mathbf{Q}$. From practical experiments we have noticed that forcing the first principal point to the center of the image prevents the minimization from falling into a local minima, and in fact

[2] Under our parameterization, $\mathbf{Q}$ will always be nonsingular

always leads to a better reconstruction. So practically we just have 4 unknown parameters, and we need only 3 images.

The non-linear minimizations are rather unstable and need to be started from a *close* initialization. Nevertheless, we are just looking for 6 or 4 parameters which is quite reasonable. Now let's see how we can get an initialization for the focal length f and $\mathbf{q} = \begin{pmatrix} q_1 & q_2 & q_3 \end{pmatrix}^T$.

2.2.1 Looking for the focal length

In this section, we want to obtain a value or at least an estimation of the focal length for any camera we have. It is important to specify that we do not suppose the different focal lengths to be the same over our camera set. Hartley has already found it was possible [Har92] and gave the algorithm to compute from two images f and f' as soon as $\mathbf{p}$ and $\mathbf{p}'$, the principal points, are known. We will come to the same conclusion but we will provide a close-form formulae which can be much more useful in order to estimate the uncertainty or the reliability of a given result. Our formulation also provides a way to understand some degenerations which Hartley did not mention.

Let's come back to the Kruppa equations. For any pair of images, if we note $\mathbf{F}$ the fundamental matrix, $\mathbf{e}'$ the epipole in the second image, respectively $\mathbf{A}$ and $\mathbf{A}'$ the internal parameters matrices of the first and second image, and $\mathbf{K}$ the matrix $\mathbf{AA}^T$, then we have with x an unknown factor [ZF96]:

$$\mathbf{FKF}^T = x^2 \left[\mathbf{e}'\right]_\times \mathbf{K}' \left[\mathbf{e}'\right]_\times^T \quad (3)$$

Under our assumptions we have:

$$\mathbf{K} = \mathbf{AA}^T = f^2\tilde{\mathbf{I}} + \mathbf{pp}^T \text{with } \tilde{\mathbf{I}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Let's take $\mathbf{f}_p = \mathbf{Fp}$, $\mathbf{e}'_p = \left[\mathbf{e}'\right]_\times \mathbf{p}'$, $\tilde{\mathbf{F}} = \mathbf{F}\tilde{\mathbf{I}}\mathbf{F}^T$ and

$$\tilde{\mathbf{E}}' = \left[\mathbf{e}'\right]_\times \tilde{\mathbf{I}} \left[\mathbf{e}'\right]_\times^{\mathbf{T}} = \begin{bmatrix} \mathbf{e}'^2_3 & 0 & -\mathbf{e}'_1\mathbf{e}'_3 \\ 0 & \mathbf{e}'^2_3 & -\mathbf{e}'_2\mathbf{e}'_3 \\ -\mathbf{e}'_1\mathbf{e}'_3 & -\mathbf{e}'_2\mathbf{e}'_3 & \mathbf{e}'^2_1 + \mathbf{e}'^2_2 \end{bmatrix} \quad (4)$$

So from (3) we have,

$$f^2\tilde{\mathbf{F}} + \mathbf{f}_p\mathbf{f}_p^T = x^2(f'^2\tilde{\mathbf{E}}' + \mathbf{e}'_p\mathbf{e}'^T_p) \quad (5)$$

Remarking that $\tilde{\mathbf{E}}'_{12} = 0$, we have from (5) using equation $[1, 2]$:

- if $\mathbf{e}'_{p_1}\mathbf{e}'_{p_2} = 0$ we simply get: $f^2\tilde{\mathbf{F}}_{12} + \mathbf{f}_{p_1}\mathbf{f}_{p_2} = 0$
- if $\mathbf{e}'_{p_1}\mathbf{e}'_{p_2} \neq 0$: $x^2 = \frac{f^2\tilde{\mathbf{F}}_{12} + \mathbf{f}_{p_1}\mathbf{f}_{p_2}}{\mathbf{e}'_{p_1}\mathbf{e}'_{p_2}}$

Thus putting x^2 back into (5), we can write:

$$\begin{aligned} &f^2f'^2\tilde{\mathbf{F}}_{12}\tilde{\mathbf{E}}' + f^2(\tilde{\mathbf{F}}_{12}\mathbf{e}'_p\mathbf{e}'^T_p - \mathbf{e}'_{p_1}\mathbf{e}'_{p_2}\tilde{\mathbf{F}}) \\ &+ f'^2\mathbf{f}_{p_1}\mathbf{f}_{p_2}\tilde{\mathbf{E}}' + \mathbf{f}_{p_1}\mathbf{f}_{p_2}\mathbf{e}'_p\mathbf{e}'^T_p - \mathbf{e}'_{p_1}\mathbf{e}'_{p_2}\mathbf{f}_p\mathbf{f}_p^T = 0 \end{aligned}$$

This is a polynomial equation of degree 2 in f^2 and f'^2 with matrix coefficients. We can rewrite it as $f^2f'^2\mathbf{A} + f^2\mathbf{B} + f'^2\mathbf{C} + \mathbf{D} = \mathbf{0}_{3\times3}$ with $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ 4, 3×3 symmetric matrices. Let's call EQ those 6 equations. EQ_{12} was already used for the parameters x. It's not too difficult to show that $\mathbf{e}_1 EQ_{11} = -\mathbf{e}_3 EQ_{13}$ and $\mathbf{e}_2 EQ_{21} = -\mathbf{e}_3 EQ_{23}$, so we have 3 equations left, the one which are on the diagonal of EQ. Now the key issue is that $\tilde{\mathbf{E}}'_{11} = \tilde{\mathbf{E}}'_{22}$ so we have $\mathbf{A}_{11} = \mathbf{A}_{22}$, $\mathbf{C}_{11} = \mathbf{C}_{22}$ and:

$$EQ_{11} - EQ_{22} = 0 \Leftrightarrow (\mathbf{B}_{11} - \mathbf{B}_{22})f^2 + \mathbf{D}_{11} - \mathbf{D}_{22} = 0 \quad (6)$$

it can be shown that:

$$\begin{aligned} &\begin{cases} \mathbf{B}_{11} = \tilde{\mathbf{F}}_{12}\mathbf{e}'_{p_1}\mathbf{e}'_{p_1} - \tilde{\mathbf{F}}_{11}\mathbf{e}'_{p_1}\mathbf{e}'_{p_2} \\ \mathbf{D}_{11} = \mathbf{f}_{p_1}\mathbf{f}_{p_2}\mathbf{e}'_{p_1}\mathbf{e}'_{p_1} - \mathbf{e}'_{p_1}\mathbf{e}'_{p_2}\mathbf{f}_{p_1}\mathbf{f}_{p_1} \end{cases} \\ \Leftrightarrow &\begin{cases} \mathbf{B}_{11} = -\mathbf{e}'_{p_1}\begin{pmatrix} f_{11} & f_{12} & 0 \end{pmatrix}\mathbf{F}^T\mathbf{p}'\mathbf{e}'_3 \\ \mathbf{D}_{11} = -\mathbf{f}_{p_1}\mathbf{e}'_{p_1}(\mathbf{p}^T\mathbf{F}\mathbf{p}')\mathbf{e}'_3 \end{cases} \end{aligned}$$

as well as $\begin{cases} \mathbf{B}_{22} = \mathbf{e}'_{p_2}\begin{pmatrix} f_{21} & f_{22} & 0 \end{pmatrix}\mathbf{F}^T\mathbf{p}'\mathbf{e}'_3 \\ \mathbf{D}_{22} = -\mathbf{f}_{p_2}\mathbf{e}'_{p_2}(\mathbf{p}^T\mathbf{F}\mathbf{p}')\mathbf{e}'_3 \end{cases}$

If $\mathbf{e}'_3 = 0$ we remark that $\mathbf{e}'_1 EQ_{11} = -\mathbf{e}'_3 EQ_{13}$ and $\mathbf{e}'_2 EQ_{21} = -\mathbf{e}'_3 EQ_{23}$. So, as EQ_{13} and EQ_{23} should be verified, we can always simplify (6) by $\mathbf{e}'_3$ and write:

$$(6) \Leftrightarrow \mathbf{e}'^T_p\tilde{\mathbf{I}}(\tilde{\mathbf{F}}f^2 + \mathbf{Fpp}^T\mathbf{F}^T)\mathbf{p}' = 0$$

Which means that in every case, we have:

$$f = \sqrt{-\frac{\mathbf{p}'^T\left[\mathbf{e}'\right]_\times \tilde{\mathbf{I}}\mathbf{Fpp}^T\mathbf{F}^T\mathbf{p}'}{\mathbf{p}'^T\left[\mathbf{e}'\right]_\times \tilde{\mathbf{I}}\mathbf{F}\tilde{\mathbf{I}}\mathbf{F}^T\mathbf{p}'}} \quad (7)$$

Which is a straight forward estimation of f that relies only on the knowledge of $\mathbf{F}$, $\mathbf{p}$ and $\mathbf{p}'$ whatever f' could be. For practical situations, many people agree that the calibration does not really depend on the principal point, and we will explain it in section 4.2. So for this estimation we simply put them both on the center of the images. We must keep in mind that this is a practical estimation, that this equation may not be well conditioned. Furthermore there may remain some degenerated cases, as when one epipole is aligned to the principal point, and when the principal points are in epipolar correspondences. For the time being, we do not know what happens under such particular situations. Here we can compute the focal length from every $\mathbf{F}_{0i}$, and we keep the reasonable ones *i.e.* the ones that belong to [300..1200].

2.2.2 Looking for q

Now we have an estimation of $\mathbf{A}_i$. We note for each i, $\mathbf{P}^i_{proj} = [\ \tilde{\mathbf{P}}_i \quad \tilde{\mathbf{t}}_i\ ]$, so we have:

$$\mathbf{P}^i_{euc} = \mathbf{P}^i_{proj}\mathbf{Q} = [\ \tilde{\mathbf{P}}_i\mathbf{A}_0 + \tilde{\mathbf{t}}_i\mathbf{q}^T \quad \tilde{\mathbf{t}}_i\ ]$$

As $\mathbf{P}^i_{euc} = [\ \mathbf{A}_i\mathbf{R}_i \quad \mathbf{A}_i\mathbf{t}_i\ ]$, we have $\tilde{\mathbf{P}}_i\mathbf{A}_0 + \tilde{\mathbf{t}}_i\mathbf{q}^T \simeq \mathbf{A}_i\mathbf{R}_i$. For an approximation we can take $\mathbf{A}_0 = \mathbf{A}_i$ and thus for any particular matrix $\mathbf{P}_{euc}$:

$$\mathbf{R}\mathbf{R}^T = \mathbf{I} \Leftrightarrow (\tilde{\mathbf{P}}\mathbf{A} + \tilde{\mathbf{t}}\mathbf{q}^T)(\tilde{\mathbf{P}}\mathbf{A} + \tilde{\mathbf{t}}\mathbf{q}^T)^T = x^2\mathbf{K} \Leftrightarrow$$
$$\tilde{\mathbf{P}}\mathbf{K}\tilde{\mathbf{P}}^T + \tilde{\mathbf{t}}\mathbf{q}^T\mathbf{A}^T\tilde{\mathbf{P}}^T + \tilde{\mathbf{P}}\mathbf{A}\mathbf{q}\tilde{\mathbf{t}}^T + \|\mathbf{q}\|^2\tilde{\mathbf{t}}\tilde{\mathbf{t}}^T = x^2\mathbf{K}$$

As $\mathbf{K}_{33} = 1$, $x^2 = \tilde{\mathbf{p}}_3^T\mathbf{K}\tilde{\mathbf{p}}_3 + 2\tilde{\mathbf{p}}_3^T\mathbf{A}\mathbf{q}\tilde{\mathbf{t}}_3 + \|\mathbf{q}\|^2\tilde{\mathbf{t}}_3^2$

Moreover we can see that the quadratic term in $\mathbf{q}$ is constant over our 5 remaining equations. So we look for i_0 and j_0 such that $i_0 \in [1..2]$, $j_0 \in [i_0..3]$ and $|\tilde{\mathbf{t}}_{i_0}\tilde{\mathbf{t}}_{j_0}^T - \tilde{\mathbf{t}}_3^2\mathbf{K}_{i_0j_0}|$ is maximum. Then we extract $\|\mathbf{q}\|^2$ from this equation and substitute its expression back into the other 4.

$$\|\mathbf{q}\|^2 = (\tilde{\mathbf{p}}_{i_0}^T\mathbf{K}\tilde{\mathbf{p}}_{j_0} - \tilde{\mathbf{p}}_3^T\mathbf{K}\tilde{\mathbf{p}}_3\mathbf{K}_{i_0j_0} + (\tilde{\mathbf{t}}_{i_0}\tilde{\mathbf{p}}_{j_0}^T + \tilde{\mathbf{t}}_{j_0}\tilde{\mathbf{p}}_{i_0}^T - 2\mathbf{K}_{i_0j_0}\tilde{\mathbf{t}}_3\tilde{\mathbf{p}}_3^T)\mathbf{A}\mathbf{q})/(\tilde{\mathbf{t}}_3^2\mathbf{K}_{i_0j_0} - \tilde{\mathbf{t}}_{i_0}\tilde{\mathbf{t}}_{j_0})$$

So,
$$\begin{aligned}&(\tilde{\mathbf{t}}_3^2\mathbf{K}_{i_0j_0} - \tilde{\mathbf{t}}_{i_0}\tilde{\mathbf{t}}_{j_0})[\tilde{\mathbf{p}}_i^T\mathbf{K}\tilde{\mathbf{p}}_j - \tilde{\mathbf{p}}_3^T\mathbf{K}\tilde{\mathbf{p}}_3\mathbf{K}_{ij} \\ &+(\tilde{\mathbf{t}}_i\tilde{\mathbf{p}}_j^T + \tilde{\mathbf{t}}_j\tilde{\mathbf{p}}_i^T - 2\mathbf{K}_{ij}\tilde{\mathbf{t}}_3\tilde{\mathbf{p}}_3^T)\mathbf{A}\mathbf{q}] + \\ &(\tilde{\mathbf{t}}_i\tilde{\mathbf{t}}_j - \tilde{\mathbf{t}}_3^2\mathbf{K}_{ij})[\tilde{\mathbf{p}}_{i_0}^T\mathbf{K}\tilde{\mathbf{p}}_{j_0} - \tilde{\mathbf{p}}_3^T\mathbf{K}\tilde{\mathbf{p}}_3\mathbf{K}_{i_0j_0} \\ &+(\tilde{\mathbf{t}}_{i_0}\tilde{\mathbf{p}}_{j_0}^T + \tilde{\mathbf{t}}_{j_0}\tilde{\mathbf{p}}_{i_0}^T - 2\mathbf{K}_{i_0j_0}\tilde{\mathbf{t}}_3\tilde{\mathbf{p}}_3^T)\mathbf{A}\mathbf{q}] = 0\end{aligned}$$

Which is of the form $\mathbf{a}_{i_0j_0ij}\mathbf{A}\mathbf{q} + b_{i_0j_0ij} = 0$, each image (excluding the first one) gives us 4 linear equations in $\mathbf{A}\mathbf{q}$. We estimate this system with a classical *least squares algorithm.*

3 Experiments

For this study we took 6 images of the INRIA's library from different locations. We tried to avoid degenerated cases, and we explicitly moved away and zoomed to change the focal lengths. We defined manually thanks to [BR97] about 60 matches within each image, estimated the internal parameters with a classical approach based on the Kruppa equations [ZF96], which gave us the meaningless internal parameters $f = 10$ and $(u_0, v_0) = (317, 277)$. Then we launched a bundle adjustment in the Euclidean space which ended

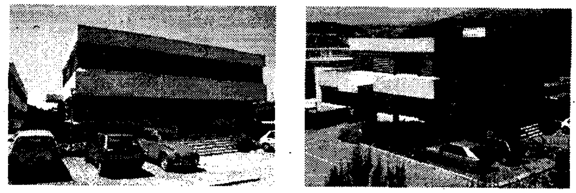

Figure 1: Images 1,5 of our sequence

up with a residual of 188 pixels, definitely weird internal parameters and obviously the reconstruction was nonsense!

Now we launched over the same data a bundle adjustment in the projective space. It ended up with a residual of 0.06 pixels, so we do think that this bundle adjustment picked up the scene's geometry in projective space.

The estimation for the focal length of the first camera using the fundamental matrices we had and (7), was:

	i_1	i_2	i_3	i_4	i_5
f	575	590	683	535	418

So we chose to manually initialize our process with $f = 575$, the estimation of $\mathbf{q}$ as describe in section 2.2.2 gave us:

$$\mathbf{q} = (\ 9.35 \quad 14.97 \quad 16.45\ )$$

Then the final minimization gave us,

$$\mathbf{Q} = \begin{bmatrix} 598.3 & 0 & 384 & 0 \\ 0 & 598.3 & 256 & 0 \\ 0 & 0 & 1 & 0 \\ 12.20 & 16.48 & 11.36 & 1 \end{bmatrix}$$

and with such a collineation we get these internal parameters:

α_u	α_v	u_0	v_0	α_u/α_v	γ
598.3	598.3	384	256	1	0
699.0	696.3	360	257	1.004	0.003
844.6	841.6	315	255	1.003	-0.0029
910.2	911.9	292	249	0.998	0.00029
1130.3	1133.7	244	240	0.997	0.0008
976.6	974.3	317	236	1.002	-0.0010

Within such a basis, we obtain a visually perfect reconstruction as shown in Figure 2. 'Visually perfect' means that from our comprehension of the scene we observe that the different angles supposedly right look pretty right, the parallel lines look parallel and the different lengths of the scene look coherent. We ended up with a calibration that looks much better than the one we could obtain before.

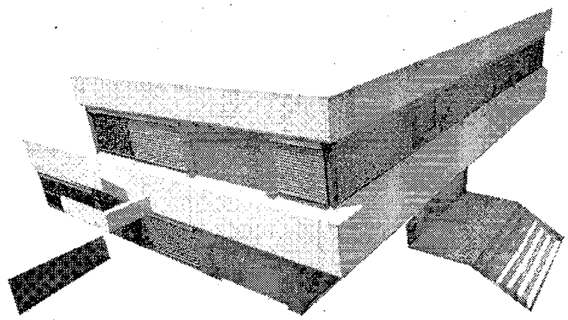

Figure 2: Rough reconstruction of INRIA's library in our new Euclidean basis. Get it at:
http://www.inria.fr/robotvis/personnel/sbougnou/Results/

4 The internal parameters estimation problem, a criticism of self-calibration

4.1 The focal length

In practical experiments, we found that the estimation of the focal length is rather unstable *e.g.* it can easily range from 500 to 800 still yielding very good residual for our set of constraints. For instance starting with $f = 650$ yields:

α_u	α_v	u_0	v_0	α_u/α_v	γ
759.1	759.1	384	256	1	0
905.1	906.5	373	275	0.998	0.0029
1136.1	1138.5	366	278	0.997	-0.0032
1242.2	1243.8	387	268	0.999	0.0011
1775.3	1766.2	331	496	1.005	0.0003
1394.1	1386.8	329	404	1.005	-0.005

In fact, if in theory internals and externals parameters are separable, the focal length and the different translations under practical situation are highly correlated. For example there is a strong ambiguity between a translation along the focal axis and a zoom. Nevertheless, if we cannot numerically raise this ambiguity, it seems that it does not affect the quality of the reconstruction. Let's understand this assumption a bit further.

From an initialization of f, our minimization gives us a collineation $\mathbf{Q}$, then starting with f' we obtain $\mathbf{Q}'$. With such collineations, we can compute the collineation $\mathbf{H}^{euc}_{euc'}$ between the two obtained Euclidean basis:

$$\mathbf{H}^{euc}_{euc'} = \mathbf{Q}'^{-1}\mathbf{Q}$$

If we note as usual:

$$\mathbf{Q} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ q_1 & q_2 & q_3 & 1 \end{bmatrix} \quad \mathbf{Q}' = \begin{bmatrix} f' & 0 & u_0 & 0 \\ 0 & f' & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ q'_1 & q'_2 & q'_3 & 1 \end{bmatrix}$$

Then we have:

$$\mathbf{H}^{euc}_{euc'} = \begin{bmatrix} \frac{f}{f'} & 0 & 0 & 0 \\ 0 & \frac{f}{f'} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{q_1 f' - q'_1 f}{f'} & \frac{q_2 f' - q'_2 f}{f'} & q_3 - q'_3 & 1 \end{bmatrix}$$

First note that the unknown transformation $\mathbf{H}^{euc}_{euc'}$ does not depend on the principal point we had chosen! This is because we decided to take $u'_0 = u_0$ and $v'_0 = v_0$, which we will discuss in the next sub-section. As one will never obtain reconstruction better than up to a scale factor, we allow similarities for our unknown transformation. This transformation $\mathbf{H}^{euc}_{euc'}$ does not look like an homothety because the homothety factor on the z-axis is not equal to the other two, and it has non-zero coefficients on its last row. This last row represents the equation of the plane at infinity of the second basis within the coordinate of the first one. Now let's see more precisely what happens to the plane at infinity under practical situations.

First we ran several minimizations changing the initialization of the focal length, then plotted the final estimation for the focal length in Figure 3. Every final

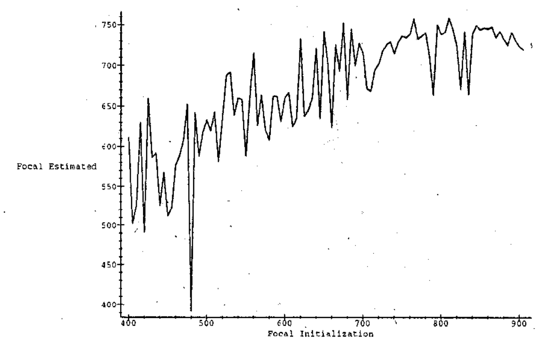

Figure 3: Final focal length estimation after minimizing our constraints from different initializations

residuals was pretty good, *i.e.* the sum of squared residuals from the final estimation of our 10 constraints was less than 5.10^{-5}.

For any two different estimation of $\mathbf{Q}$ we computed as described above the related transformation $\mathbf{H}^{euc}_{euc'\,ij}$. Then we computed the distance of any point of our scene to the plane at infinity. The key issue here is that for any case, even the worst, we noticed any such distance to be almost constant; which means that, wherever the plane at infinity is, it is sufficiently far away from our scene for any such distances to be considered constant. So if we call $\overline{d}_\infty$ this average distance, we can think of $\mathbf{H}^{euc}_{euc'}$ as being equivalent to:

$$\mathbf{H}^{euc}_{euc'} \equiv \begin{bmatrix} \frac{f}{f'} & 0 & 0 & 0 \\ 0 & \frac{f}{f'} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \overline{d}_\infty \end{bmatrix}$$

which is almost equivalent to an homothety; in fact it is an anisotropic homothety. In other words, the choice of the plane at infinity induces totally different internal parameters, but does not noticeably corrupt the reconstruction of our scene.

According to Figure 3, we can say that there is no way to compute the real focal length of the cameras. To understand the effect of this ambiguity on our reconstruction, let's see what happens to right angles. In fact, as shown on Figure 4, depending on the orientation of the right angles within the scene, they will be visually affected or not by this unknown anisotropic homothety.

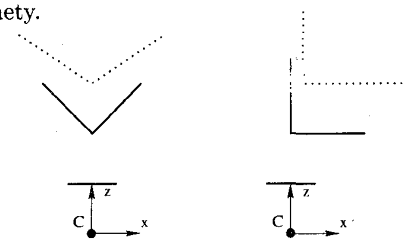

Figure 4: What may happened to right angle under our unknown anisotropic homothety

We could not look at every reconstruction for obvious time reasons, but we did examine some of them, and they were really good-looking. The fact is that the recovery of the internal parameters is not really important for the visual quality of the reconstruction. Minimizing our constraints may lead to different focal lengths that always produce really acceptable reconstructions. Now let's see why in practical situations we do not care about the principal point.

4.2 The principal point

Let's examine in more detail the error we may have made forcing the first principal point to the center of the first image. Suppose we have in fact:

$$\mathbf{Q} = \begin{bmatrix} f & 0 & u_0 & 0 \\ 0 & f & v_0 & 0 \\ 0 & 0 & 1 & 0 \\ q_1 & q_2 & q_3 & 1 \end{bmatrix} \quad \mathbf{Q}' = \begin{bmatrix} f' & 0 & u'_0 & 0 \\ 0 & f' & v'_0 & 0 \\ 0 & 0 & 1 & 0 \\ q'_1 & q'_2 & q'_3 & 1 \end{bmatrix}$$

Then we have: $\mathbf{H}^{euc}_{euc'} = \mathbf{Q'}^{-1}\mathbf{Q} =$

$$\begin{bmatrix} \frac{f}{f'} & 0 & \frac{u_0 - u'_0}{f'} & 0 \\ 0 & \frac{f}{f'} & \frac{v_0 - v'_0}{f'} & 0 \\ 0 & 0 & 1 & 0 \\ \frac{q_1 f' - q'_1 f}{f'} & \frac{q_2 f' - q'_2 f}{f'} & q_3 - q'_3 + \frac{q'_1(u'_0 - u_0) + q'_2(v'_0 - v_0)}{f'} & 1 \end{bmatrix}$$

At this point there are two things we can say:

- looking at the upper-left 3×3 sub-matrix of $\mathbf{H}^{euc}_{euc'}$, we see that the coefficients $\frac{u_0 - u'_0}{f'}$, $\frac{v_0 - v'_0}{f'}$ should be compare to 1. Then, an numerical analysis tells us that if we think of f as being close to 700, an error for (u_0, v_0) up to 50 pixels means $\frac{|u_0 - u'_0|}{f'} < 0.072$ which is negligible compared to 1,
- Moreover, as soon as one allows a translation for the camera's displacement, the deviation imposed by an inaccurate estimation of the principal point, can be mainly interpreted as a translation. In fact, the key issue is that for many situation the scene's viewable points are confined in z-coordinate due to matching purposes. So if we note $\bar{z}$ the average z for our set of 3D points, $\frac{|z-\bar{z}|}{\bar{z}} \ll 1$. Thus we have:

$$\forall M, \mathbf{H}^{euc}_{euc'} M =$$

$$\begin{pmatrix} \frac{f}{f'}x + \frac{u_0 - u'_0}{f'} z \\ \frac{f}{f'}y + \frac{v_0 - v'_0}{f'} z \\ z \\ \pi'^T_\infty M \end{pmatrix} = \begin{pmatrix} \frac{f}{f'}x + \frac{u_0 - u'_0}{f'}\frac{z-\bar{z}}{z} z + \frac{u_0 - u'_0}{f'}\bar{z} \\ \frac{f}{f'}y + \frac{v_0 - v'_0}{f'}\frac{z-\bar{z}}{z} z + \frac{v_0 - v'_0}{f'}\bar{z} \\ \frac{f}{f'}z + \frac{f'-f}{f'}\frac{z-\bar{z}}{z} z + \frac{f'-f}{f'}\bar{z} \\ \pi'^T_\infty M \end{pmatrix}$$

So now, the error we committed for the principal point and the focal length estimations can be interpreted as the translation:

$$\mathbf{t} = \begin{pmatrix} \frac{u_0 - u'_0}{f'}\bar{z} & \frac{v_0 - v'_0}{f'}\bar{z} & \frac{f'-f}{f'}\bar{z}\overline{d}_\infty \end{pmatrix}$$

And we can consider our scene to be defined up to:

$$\mathbf{H}^{euc}_{euc'} \equiv \begin{bmatrix} \frac{f}{f'} & 0 & 0 & \frac{u_0 - u'_0}{f'}\bar{z} \\ 0 & \frac{f}{f'} & 0 & \frac{v_0 - v'_0}{f'}\bar{z} \\ 0 & 0 & \frac{f}{f'} & \frac{f'-f}{f'}\bar{z} \\ 0 & 0 & 0 & \overline{d}_\infty \end{bmatrix}$$

which is a perfectly allowed similarity.

In our experiment, we had for any point in the reference frame of the first camera $\frac{|z-\bar{z}|}{\bar{z}} < 0.2$. Such considerations means that the coefficient we neglected *e.g.* $\frac{u_0 - u'_0}{f'}\frac{|z-\bar{z}|}{\bar{z}}$, allowing an error of 50 pixels for the principal point, weighted $0.072 * 0.2 = 1.44\%$ compares to 1!

Thus from a practical point of view, we can say that there is no way to estimate the principal point when:

• the focal length is wide. This explains the fact that when the focal length tends toward infinity, the principal point is not defined anymore;

• the z-coordinate of the set of points you are looking at, are far from the camera and confined into narrow boundaries, which is fortunately or unfortunately often the case.

In such situations, any given values can be interpreted as a translation. Thus forcing the principal point to the center of the image during the minimization process prevents it from coming up to nonsense values, and the resulting error can be considered to be less than a few percent for the reconstruction. Meanwhile, the fact that the coordinates of the principal point does not affect the quality of the reconstruction does not mean that one should constraint it to be constant over the set of images. Those constrains are too weak for a minimization point of view, and may result in corrupting or biasing some others. In fact, according to our projection matrices, it moves around within the images but we do not care about an accurate estimation of its position.

5 Conclusion

Starting from a projective set of projection matrices and imposing the aspect ratio to be 1 and the skew to be 0 allow to get the Euclidean scene's structure back. The described method needs at least 3 images without assuming anything on the internal parameters, nor on the displacements. For numerical reasons, we found that the structure of the scene can be obtained up to an anisotropic homothety, close from an allowed homothety, resulting in the difficulty of estimating the focal length of the reference camera. We have also shown theoretically why the estimation of the principal point is so unstable but fortunately not at all important for the quality of the 3D reconstruction. Even if the obtained calibration may remain insufficient for vision-guided tasks because of the ambiguity on the translation, it is perfectly suitable in order to get a reconstructed model.

References

[AZB94] Martin Armstrong, Andrew Zisserman, and Paul Beardsley. Euclidean structure from uncalibrated images. In Edwin Hancock, York, UK, September 1994. BMVA PressHancock, pages 508–518.

[BR97] Sylvain Bougnoux and Luc Robert. TotalCalib: a fast and reliable system for off-line calibration of images sequences. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, June 1997. The Demo Session.

[Fau93] Olivier Faugeras. *Three-Dimensional Computer Vision: a Geometric Viewpoint.* MIT Press, 1993.

[Fau94] Olivier Faugeras. Stratification of 3-D vision: projective, affine, and metric representations. *Journal of the Optical Society of America A*, 1994.

[GB92] Armin Gruen and Horst A. Beyer. System calibration through self-calibration. In *Proceedings of the Workshop on Calibration and Orientation of Cameras in Computer Vision*, Washington D.C., August 1992.

[HÅ97] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Comp. Vision and Pattern Rec.*, pages 438–443. IEEE Computer Society Press, 1997.

[Har92] Richard Hartley. Calibration of cameras using the essential matrix. In *Proceedings of the ARPA Image Understanding Workshop*, pages 911–915. Defense Advanced Research Projects Agency, Morgan Kaufmann Publishers, Inc., 1992.

[Har94] Richard I. Hartley. An algorithm for self calibration from several views. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 908–912, Seattle, WA, June 1994. IEEE.

[Lav96] Stéphane Laveau. *Géométrie d'un système de N caméras. Théorie, estimation et applications.* PhD thesis, École Polytechnique, May 96.

[LF97] Q.-T. Luong and O. Faugeras. Self-calibration of a moving camera from point correspondences and fundamental matrices. *The International Journal of Computer Vision*, 22(3):261–289, 1997.

[MF92] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *The International Journal of Computer Vision*, 8(2):123–152, August 1992.

[PVG97] M. Pollefeys and L. Van Gool. A stratified approach to metric self-calibration. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 407–412.

[Ste95] Gideon P. Stein. Accurate internal camera calibration using rotation with analysis of sources of error. In *Proceedings of the 5th International Conference on Computer Vision*, Boston, MA, June 1995. IEEE Computer Society Press.

[Tri97] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pages 609–614.

[VF96] Thierry Viéville and Olivier Faugeras. The first order expansion of motion equations in the uncalibrated case. *CVGIP: Image Understanding*, 64(1):128–146, July 1996.

[ZF96] Cyril Zeller and Olivier Faugeras. Camera self-calibration from video sequences: the Kruppa equations revisited. Research Report 2793, INRIA, February 1996.

[ZLF96] Zhengyou Zhang, Quang-Tuan Luong, and Olivier Faugeras. Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction. *IEEE Transactions on Robotics and Automation*, 12(1):103–113, February 1996.