Euclidean self-calibration via the modulus constraint

Marc Pollefeys, Luc Van Gool and André Oosterlinck
Katholieke Universiteit Leuven, E.S.A.T. / MI2
Marc.Pollefeys@esat.kuleuven.ac.be

Abstract

To obtain a Euclidean reconstruction from images the cameras have to be calibrated. In the last years different approaches have been proposed to avoid explicit calibration. The problem with these methods is that a lot of parameters have to be retrieved at once. Because of the non-linearity of the equations this is not an easy task and the methods often fail to converge.

In this paper a stratified approach is proposed which allows to first retrieve the affine calibration of the camera using the *modulus constraint*. Having the affine calibration it is easy to upgrade to Euclidean. The important advantage of this method is that only three parameters have to be evaluated at first. From a practical point of view, the major gain is that an affine reconstruction is obtained from arbitrary sequences of views, whereas so far affine reconstruction has been based on pairs of views with a pure translation in between.

1 Introduction

Several researchers have shown the possibility of calibrating a camera from correspondences between several views of the same scene. These methods are based on the rigidity of the scene and on the constancy of the internal parameters. Maybank [8] and Faugeras [3] extracted two quadratic constraints in the five unknown internal parameters for each pair of views. Solving these equations needs high accuracy computations. The number of potential solutions grows exponentially with the number of views which makes this method intractable for a large number of views. An alternative method was proposed by Hartley [6] who solved for the eight unknowns of the affine and Euclidean calibration at once. This method was more robust and worked better with a large number of views. The disadvantage is that the method fails to converge if the initialization isn't near to the final solution which isn't easy to guarantee in an eight parameter space [a].

This problem prompted a stratified approach, where an affine reconstruction is obtained first and used as the initialization towards Euclidean recon-

[a] certainly not for the affine calibration parameters which are not fixed for a camera unlike the internal camera parameters

struction. Such a method has been proposed by Armstrong *et al*[1] based on the work of Moons *et al*[9].

Also in this paper a stratified approach is given which first retrieves the affine calibration of the cameras using the *modulus constraint* and then uses additional constraints to upgrade the calibration to Euclidean. The advantage of this method compared to Hartley's is that the non-linear minimization only takes place in a three dimensional parameter space which means in practice that we can always converge to the optimal solution. Armstrong's method requires a pure translation which might be difficult to achieve with a hand-held camera, for instance. Allowing general motions is thus one of the main advantages of the method proposed in this paper.

2 Euclidean, affine and projective cameras

In this paper a pinhole camera model will be used. Central projection forms an image on a light-sensitive plane. The following equation expresses the relation between image points and world points.

$$\lambda_{ik} m_{ik} = \mathbf{P}_k M_i \tag{1}$$

Here $\mathbf{P}_k$ is a 3×4 camera matrix, m_{ik} and M_i are column vectors containing the homogeneous coordinates of the image points resp. world points, λ_{ik} expresses the equivalence up to a scale factor.

Now the camera model of Equation (1) will be specialized to the case where the Euclidean calibration is known. From there the affine and projective case and the relations between all these strata will be highlighted.

The projection matrices of the same Euclidean camera for different views can be represented as follows

$$\begin{aligned} \mathbf{P}_{E1} &= \lambda_1 \mathbf{K}\,[\,\mathbf{I}\,|\,\mathbf{0}\,] \\ \mathbf{P}_{Ek} &= \lambda_k \mathbf{K}\,[\mathbf{R}_k | - \mathbf{R}_k \mathbf{t}_k] \end{aligned} \tag{2}$$

with $\mathbf{K}$ the calibration matrix, with $\mathbf{R}_k$ and $\mathbf{t}_k$ representing the orientation and position with respect to the first camera and with λ_k a random scale factor because a projection matrix is only defined up to scale. $\mathbf{K}$ is an upper triangular matrix of the following form:

$$\mathbf{K} = \begin{bmatrix} r_x^{-1} & s & u_x \\ & r_y^{-1} & u_y \\ & & 1 \end{bmatrix} \tag{3}$$

with r_x and r_y the pixel dimensions, $u = (u_x, u_y)$ the principal point and s a skew factor.

Having only an affine calibration of the cameras means that one doesn't know what the calibration matrix $\mathbf{K}$ is. Therefore the most evident choice is to take $\mathbf{P}_{A1} = [\mathbf{I}|\mathbf{0}]$. Which means that the Affine projection matrices can be obtained from the Euclidean ones by the following transformation

$$\mathbf{T}_{AE} = \begin{bmatrix} \mathbf{K}^{-1} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} . \tag{4}$$

This yields the following affine projection matrices

$$\mathbf{P}_{A1} = \lambda_1 [\mathbf{I} \,|\, \mathbf{0}] \tag{5}$$

$$\mathbf{P}_{Ak} = \lambda_k \left[\mathbf{K}\mathbf{R}_k\mathbf{K}^{-1} | -\mathbf{K}\mathbf{R}_k\mathbf{t}_k\right] . \tag{6}$$

From these equations one can notice that the left 3×3 part of the matrices $\mathbf{P}_{Ak}$ –which will be called $\tilde{\mathbf{P}}_{Ak}$ further on– is conjugated to the scaled rotation matrix $\lambda_k \mathbf{R}_k$ and hence all eigenvalues must have equal moduli ($=\lambda_k$). This is the *modulus constraint* which will be used further on.

What can be retrieved in practice are projection matrices which are defined up to a projective transformation[2]. Choosing $\mathbf{P}_{P1} = [\mathbf{I}|\mathbf{0}]$ does not completely determine the other projection matrices. There are still 4 degrees of freedom left[b]. Projective projection matrices of the following form can be assumed:

$$\mathbf{P}_{P1} = \lambda_1 [\mathbf{I} \,|\, \mathbf{0}] \tag{7}$$

$$\mathbf{P}_{Pk} = \lambda_k \left[p_{kij} | p_{ki4}\right] . \tag{8}$$

To go from projective camera matrices to affine camera matrices the following transformation must exist:

$$\mathbf{P}_{Ak} = \mathbf{P}_{Pk}\mathbf{T}_{AP} \text{ with } \mathbf{T}_{AP} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{a} & a_4 \end{bmatrix} . \tag{9}$$

In the above equation $\mathbf{a} = [a_1 \, a_2 \, a_3]$ is a vector containing 3 parameters and a_4 is a parameter that can arbitrarily be put to 1, thereby fixing the absolute scale factor of the 3D scene[c].

[b] A projective transformation has 15 degrees of freedom. Fixing one camera projection matrix up to scale reduces the number of degrees of freedom by 11.

[c] This scale factor can never be retrieved in a fully uncalibrated environment. There is no difference in seeing small objects from nearby and big objects from far away, which is the reason why reconstructions are always up to a similarity transformation.

3 The modulus constraints for affine calibration

In this section it will be demonstrated how projective cameras can be upgraded to affine cameras by the use of the *modulus constraint.* The projective camera matrices can be retrieved using methods described in the literature (see for example Rothwell *et al*[11]). These are related to the affine ones by Equation (9). On the other hand the affine cameras are also related to the Euclidean ones through Equation (6), leading to the *modulus constraint.* This means that the *modulus constraint* must be valid for the affine camera matrices given in Equation (9) and thus can be used to determine $\mathbf{T}_{AP}$. To make the constraint explicit we write down the characteristic equation of $\tilde{\mathbf{P}}_{Ak}$:

$$\det\left(\tilde{\mathbf{P}}_{Ak} - \lambda\mathbf{I}\right) = a\lambda^3 + b\lambda^2 + c\lambda + d = 0 \ . \tag{10}$$

In the previous equation a, b, c, d represent first order polynomials in a_1, a_2 and a_3. The *modulus constraint* imposes that the roots of Equation (10) $|\lambda_1| = |\lambda_2| = |\lambda_3| (= \lambda_k)$. This constraint is not easy to impose, but the following constraint can be derived from it (see Appendix A):

$$ac^3 = b^3 d \tag{11}$$

Filling in a, b, c, d in Equation (11), one obtains a 4^{th} order polynomial equation in a_1, a_2, a_3. In fact one gets such a constraint for any camera except the first (reference) camera. The unknowns a_1, a_2, a_3 being the same for all cameras, one can find a finite number of solutions for four cameras. For more one will in general only have one solution. A solution to these equations can for example be found by using a Levenberg-Marquardt algorithm. In practice once a solution is found it can be checked for the *modulus constraint*, which is more stringent than Equation (11). Most often this yields only one possible solution even for only four views.

Having a_1, a_2, a_3 it is easy to bring everything in an affine framework. Affine projection matrices can directly be obtained from Equation (9). Points from the projective reconstruction can be transformed to the affine frame by using $\mathbf{T}_{AP}$. Finally the infinity homography $\mathbf{H}_{\infty 1k}$ which indicates the transformation of the plane at infinity from the first view to the k^{th}, can be retrieved as follow:

$$\mathbf{H}_{\infty 1k} = \mathbf{P}_{Pk} \begin{bmatrix} \mathbf{I} \\ \mathbf{a} \end{bmatrix} \ . \tag{12}$$

4 Euclidean calibration from affine

To upgrade the reconstruction to Euclidean the camera calibration matrix $\mathbf{K}$ is needed. This is equivalent to knowing the image $\mathbf{B}$ of the dual of the absolute conic, since $\mathbf{B} = \mathbf{K}\mathbf{K}^\top$. For every image k the following constraint must be valid:

$$\kappa_{1k}\mathbf{B} = \mathbf{H}_{1k\infty}\mathbf{B}\mathbf{H}_{1k\infty}^\top \tag{13}$$

with $\mathbf{H}_{1k\infty}$ the infinity homography between the two images. Scaling $\mathbf{H}_{1k\infty}$ (which is known up to scale from the affine calibration) to obtain $\det \mathbf{H}_{1k\infty} = 1$ forces $\kappa_{1k} = 1$ and yields a set of linear equations in the coefficients of $\mathbf{B}$. $\mathbf{K}$ can be obtained from $\mathbf{B}$ by cholesky factorization.

5 Degenerate cases

One has to pay attention to degenerate cases because they often occur in practice. As a first example take translational motions which immediately yield an affine reconstruction, but only this kind of motion will never allow the retrieval of the internal parameters of a camera. Intuitively one can see that e.g. some comparison between measurements taken along the x- and the y-axis is necessary to determine the aspect ratio.

Now we will give a more strict proof of degeneracy for planar motion. To upgrade a calibration from affine to Euclidean one uses the fact that the absolute conic is a fixed entity of the plane at infinity. It turns out that it is more practical to work with the dual of the absolute conic $\mathbf{B} = \mathbf{K}\mathbf{K}^\top$. Imposing that the absolute conic stays put in spite of the motion is equivalent with the following constraint:

$$\mathbf{B} = \mathbf{H}_\infty\mathbf{B}\mathbf{H}_\infty^\top \tag{14}$$

For a planar movement these equations will not yield a unique solution (up to scale), but one will be left with at least a one parameter family of solutions. A planar motion means that any rotation occurring will have a rotation axis perpendicular on the plane of motion. For all these motions the same point at infinity will stay put (*i.e.* $\mathbf{H}_\infty p_\infty = \lambda p_\infty$, with p_∞ the vanishing point of the rotation axes). This means that also $p_\infty p_\infty^\top$ will be a solution of Equation (14)

$$p_\infty p_\infty^\top = \mathbf{H}_\infty p_\infty p_\infty^\top \mathbf{H}_\infty^\top \tag{15}$$

and hence any linear combination of $\mathbf{B}$ and $p_\infty p_\infty^\top$ will satisfy Equation (14). Hence, one will not be able to recover $\mathbf{B}$ uniquely. Notice that pure translation keeps the whole plane at infinity constant and hence Equation (14) is not a constraint anymore ($\mathbf{H}_\infty = \mathbf{I}$).

In conclusion one can say that attention has to be paid to the choice of image sequences which are used for self-calibration. It is important that the motion is "rich" enough to allow the identification of the internal camera parameters. For example the sequence used by Hartley [6] was degenerated. One can easily see from the images that the only motion occuring is a rotation around a central axis. Sometimes it is possible to cope with these degeneracies by using additional constraint like the orthogonality of the camera axes or a known aspect ratio. However, the orthogonality often appears to be a degenerated constraint while the aspect ratio is often not accurately known before calibration.

6 Experiments

In this section some results of self-calibration on synthetic data are presented. Results are compared with Hartley's [6]. The method is shown to work properly in the presence of realistic amounts of noise. Two experiments will be presented here. One with a small number of views and one with a larger number of views.

For both experiments the scene consisted of 50 points randomly scattered in a sphere of radius 1 unit. The cameras were given random orientations and were placed at varying distances from the center of the sphere at a mean distance from the center of 2.5 units with a standard deviation of 0.25 units. They were placed in such a way that the principal rays of the cameras passed through randomly selected points on a sphere of radius 0.1 units. The calibration matrix was given a known value to be able to asses the quality of the calibration afterwards. Normal noise with different standard deviations was added to the image projections of the scene points to analyze the robustness of the method to noise. This experimental setup is the same as the one used by Hartley [6] with 15 views. To ease the comparison the same layout was used for the results.

The first experiment was carried out on 4 views [d]. The results can be seen in Table 1. For the meaning of the parameters the reader is referred to Section 2. The first line gives the exact values, subsequent lines give the results obtained with different levels of noise. One can see that even for serious amounts of noise qualitatively good results can be obtained. These can not immediately be compared to Hartley's because he used another experimental setup and only 3 views, but still the degradation of the calibration seems to be much smaller than with his method for a small number of views. Because of

[d] 4 views being the minimum to retrieve the Euclidean calibration using this technique

the low dimensionality of the parameter space convergence was easily obtained without the need for any prior knowledge that would allow to start near the final solution.

Noise	u_x	u_y	r_y^{-1}	$skew$	r_x^{-1}/r_y^{-1}
–	500.00	400.00	1000.00	-5.00	0.9000
0.0	494.86	402.08	1035.15	-16.11	0.8889
0.5	490.64	402.16	1058.26	-23.30	0.8812
1.0	474.98	403.35	1152.03	-59.65	0.8491
2.0	458.27	410.47	1180.29	-49.99	0.8372
4.0	438.56	396.44	1360.70	-97.95	0.7856

Table 1: Calibration results for 4 views

The second experiment is the same as the previous one, but carried out on 15 views. Having a large number of views gives a lot of redundancy which allows a more precise calibration. This can be seen from the results of Table 2 which are significantly better than the ones from Table 1. The results are comparable with Hartley's although somewhat less precise for higher levels of noise. This is probably due to the fact that at the moment only linear methods were used to compute the projective reconstruction.

Noise	u_x	u_y	r_y^{-1}	$skew$	r_x^{-1}/r_y^{-1}
0.0	500.00	400.00	1000.00	-5.00	0.9000
0.5	502.01	401.14	1000.89	-5.10	0.9000
1.0	499.33	398.25	997.88	-5.74	0.8992
2.0	501.64	397.79	978.04	1.37	0.9044
4.0	495.15	410.18	960.31	-8.24	0.8902

Table 2: Calibration results for 15 views

7 Conclusion and further work

In this paper the *modulus constraint* was proposed as a new constraint for self-calibration. An important result is the ability to obtain an affine calibration from a single moving camera undergoing general motion (i.e. not restricted to pure translation as in Moons *et al*[9]). From there on it is easy to get a Euclidean reconstruction. In contrast to the methods of Maybank and Faugeras [8] and Hartley [6], this method has the advantage of requiring a non-linear optimization in only three variables which reduces convergence problems. Results seems to be better than Hartley's on short sequences and comparable on longer ones. Another possibility offered by the *modulus constraint* is Euclidean calibration in the presence of a varying focal length. This alternative use of the *modulus constraint* was worked out in Pollefeys *et al*[10].

A short discussion of some degenerated cases for the Euclidean calibration problem was given. It seems that if a sequence is restricted to planar motion no complete Euclidean calibration is possible using the standard constraints, things become worse if the motion is restricted to translation. It is important to be aware of these problems when one uses self-calibration methods.

Some further work is required to get a more robust implementation of the methods presented in this paper. Better results can be expected by using non-linear refinement of the projective cameras which were used as a starting point for the presented methods. Also a combination of the different methods for self-calibration could help.

A further ideas is to exploit the special form of $ac^3 = b^3 d$ (Eq. (11)) to ease convergence by even further reducing the dimension of the non-linear optimization problem. Indeed, $a = -1$ being trivial, b, c, d (from one of the constraints) can be substituted to a_1, a_2, a_3 yielding a trivial equation $d = -c^3 b^{-3}$ leaving only b and c as unknowns. It is also needed to investigate if the *modulus constraint* could yield an affine calibration using only three images.

Acknowledgement

Marc Pollefeys acknowledges a specialisation grant from the Flemish Institute for Scientific Research in Industry (IWT). Financial support from the EU ACTS project AC074 'VANGUARD' and from the IUAP-50 project of the Belgian OSTC is also gratefully acknowledged.

A constraint on a, b, c, d

The roots of Equation (10) must obey $|\lambda_1| = |\lambda_2| = |\lambda_3|$. A general condition on a, b, c, d for this to hold is derived next. A third order polynomial can be writen as follows.

$$a\lambda^3 + b\lambda^2 + c\lambda + d = a(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3) \tag{16}$$

From Equation (16) the following relations follow:

$$\lambda_1 + \lambda_2 + \lambda_3 = -\frac{b}{a} \tag{17}$$

$$\lambda_1(\lambda_2 + \lambda_3) + \lambda_2\lambda_3 = \frac{c}{a} \tag{18}$$

$$\lambda_1\lambda_2\lambda_3 \quad = \quad -\frac{d}{a} \tag{19}$$

We want to derive a necessary condition for $|\lambda_1| = |\lambda_2| = |\lambda_3|$. If we choose λ_1 to be real (λ_2 and λ_3 can be either real or complex), the following equivalence must be true.

$$\lambda_1^2 = \lambda_2\lambda_3 \tag{20}$$

Rewriting (18) using (17) and (20) yields

$$\lambda_1(-\frac{b}{a} - \lambda_1) + \lambda_1^2 = \frac{c}{a} \tag{21}$$

or

$$\lambda_1 = -\frac{c}{b} \tag{22}$$

substituting (20) in (19) implies

$$\lambda_1^3 = -\frac{d}{a} \tag{23}$$

Eliminating λ_1 from the Equations (22) and (23) gives a necessary condition that is only depending on a, b, c, d.

$$ac^3 = b^3 d \tag{24}$$

In Equation (24) we can substitute the equations for a, b, c, d (which are found by substituting Equation (9) in Equation (10)), yielding a 4^{th} order polynomial in a_1, a_2, a_3.

1. M. Armstrong, A. Zisserman and P. Beardsley, Euclidean structure from uncalibrated images, *Proc. BMVC'94*, 1994,
2. O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, *Proc. ECCV'92*, pp.563-578, 1992.
3. O. Faugeras, Q.-T. Luong and S. Maybank. Camera self-calibration: Theory and experiments, *Proc. ECCV'92*, pp.321-334, 1992.
4. R. Deriche, Z. Zhang, Q.-T. Luong and O. Faugeras, Robust Recovery of the Epipolar Geometry for an Uncalibrated Stereo Rig *Proc. ECCV'94*, pp.567-576, 1994.
5. R. Hartley, Estimation of relative camera positions for uncalibrated cameras, *Proc. ECCV'92*, pp.579-587, 1992.
6. R. Hartley, Euclidean reconstruction from uncalibrated views, in: J.L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of invariance in Computer Vision*, Lecture Notes in Computer Science **825**, pp. 237–256, Springer, 1994.

7. Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. *Proc. ECCV'94*, pp.589-597, 1994.
8. S. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8:2:123-151, 1992.
9. T. Moons, L. Van Gool, M. Van Diest and E. Pauwels, Affine reconstruction from perspective image pairs, Proc. Workshop on Applications of Invariance in Computer Vision II, pp.249-266, 1993.
10. M. Pollefeys, L. Van Gool and M. Proesmans, Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths, *ECCV'96*, Lecture Notes in Computer Science, Springer-Verlag, 1996.
11. C. Rothwell, G. Csurka and O.D. Faugeras, A comparison of projective reconstruction methods for pairs of views, *Proc. ICCV'95*, pp.932-937, 1995.