The modulus constraint: a new constraint for self-calibration

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 recent years different approaches have been proposed to avoid explicit calibration. The problem with these methods is that several 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.*

A short illustration of another application is also given. Once the affine calibration is known, the constraint can be used to retrieve the Euclidean calibration in the presence of a variable focal length.

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. Faugeras *et al* [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. Recently Zeller *et al* [14] proposed a more robust method to solve these constraints. Heyden [6] came up with a variant of this approach. An alternative method was proposed by Hartley [5] who solved for the eight unknowns of the affine and Euclidean calibration at once. This method worked better with a large number of views but fails to converge if the initialization is not near to the final solution which is not easy to guarantee in an eight parameter space.

This problem prompted a stratified approach, where an affine reconstruction is obtained first and used as the initialization towards Euclidean reconstruction. Such a method has been proposed by Armstrong *et al* [1] based on the work of Moons *et al* [8].

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.

The *modulus constraint* can also be used in another way. Having the affine calibration it is possible to retrieve a varying parameter. The method to cope with a varying focal length is briefly outlined in this article (for details see [9]).

2 Euclidean, affine and projective cameras

In this paper a pinhole camera model will be used. The following equation expresses the relation between

*IWT fellow (Flemish Institute for the Promotion of Scientific-Technological Research in Industry) 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 (see for example [5]).

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 $\mathbf{P}_{A1} = [\mathbf{I}|\mathbf{0}]$. This determines all 12 degrees of freedom of the affine transformation except for a global scale factor (which will appear as σ in the following equations). The affine projection matrices can be obtained from the Euclidean ones by the following transformation:

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

This yields the following affine projection matrices

$$\begin{aligned} \mathbf{P}_{A1} &= \lambda_1 [\mathbf{I} \,|\, \mathbf{0}] \\ \mathbf{P}_{Ak} &= \lambda_k \left[\mathbf{K}\mathbf{R}_k\mathbf{K}^{-1} | -\sigma \mathbf{K}\mathbf{R}_k \mathbf{t}_k\right] \ . \end{aligned} \tag{5}$$

From these equations one can notice that the left 3×3 part of the matrices $\mathbf{P}_{Ak}$ –which is the infinity homography $\mathbf{H}_{1k\infty}$ (see [1])– 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 by a weak calibration are projection matrices which are defined up to a projective transformation [12, 2, 4, 13]. Choosing $\mathbf{P}_{P1} = [\mathbf{I}|\mathbf{0}]$ only fixes 11 of the 15 degrees of freedom of the projective transformations. Besides the global scale factor σ three more parameters are free ($\mathbf{a} = [a_1\, a_2\, a_3]$ in the following equations). To go from projective camera matrices to affine camera matrices the following transformation can be used:

$$\mathbf{P}_{Ak} = \mathbf{P}_{Pk} \mathbf{T}_{AP} \text{ with } \mathbf{T}_{AP} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{a} & 1 \end{bmatrix} \ . \tag{6}$$

Such a transformation boils down to picking a different plane as the plane at infinity. The infinity homography $\mathbf{H}_{\infty 1k}$ can be retrieved as follow:

$$\mathbf{H}_{1k\infty} = \mathbf{P}_{Pk} \begin{bmatrix} \mathbf{I} \\ \mathbf{a} \end{bmatrix} \ . \tag{7}$$

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 [12]). These are related to the affine ones by Equation (6). On the other hand the affine cameras are also related to the Euclidean ones through Equation (5), leading to the *modulus constraint*. This means that the *modulus constraint* must be valid for the affine camera matrices given in Equation (6) and thus can be used to determine $\mathbf{a}$, i.e. to position the plane at infinity. To make the constraint explicit we write down the characteristic equation of $\mathbf{H}_{1k\infty}$:

$$\det(\mathbf{H}_{1k\infty} - \lambda \mathbf{I}) = a\lambda^3 + b\lambda^2 + c\lambda + d = 0 \ . \tag{8}$$

In the previous equation a, b, c, d are first order polynomials in a_1, a_2 and a_3. The *modulus constraint* imposes that the roots of Equation (8) $|\lambda_1| = |\lambda_2| = |\lambda_3| (= \lambda_k)$. This constraint is not easy to impose, but the following constraint can be derived from it:

$$ac^3 = b^3 d \tag{9}$$

Filling in a, b, c, d in Equation (9), 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 (9). Most often this yields only one possible solution even for only four views.

4 Euclidean calibration from affine

To upgrade the reconstruction to Euclidean the camera calibration matrix $\mathbf{K}$ is needed. Euclidean self-calibration is mostly obtained by localizing $\mathbf{\Omega}_{im}^{-1} = \mathbf{K}\mathbf{K}^\top$, the dual of the image of the absolute conic [5]. The absolute conic lays in the plane at infinity and is fixed under a rigid motion. These characteristics yields the following constraints for $\mathbf{\Omega}_{im}^{-1}$:

$$\mathbf{\Omega}_{im}^{-1} = \mathbf{H}_{1k\infty}\mathbf{\Omega}_{im}^{-1}\mathbf{H}_{1k\infty}^\top \qquad (10)$$

To avoid scale factors in the above equation $\mathbf{H}_{1k\infty}$ should be scaled to obtain $\det \mathbf{H}_{1k\infty} = 1$. $\mathbf{K}$ can be obtained from $\mathbf{\Omega}_{im}^{-1}$ by Cholesky factorization.

5 Another application: Euclidean calibration with a variable focal length

The *modulus constraint* can also be used for other purposes than affine calibration. A more complete description of this method can be found in [9]. The constraint depends on two conditions: *the affine calibration* and *the constancy of the internal parameters.* For each view except the first we get a valid constraint. This means that instead of "spending" the constraint on solving for affine calibration one can in the traditional scheme –where such calibration amounts from translation between the first two views– use the constraint to retrieve one changing parameter for each supplementary view. The most practical application is to allow the focal length to vary.

The first step is to model the effect of changes in focal length. These changes are relatively well described by scaling the image around the principal point u which can be expressed as follow:

$$m_{f\,ik} = \mathbf{K}_f m_{ik} \text{ with } \mathbf{K}_f = \begin{bmatrix} 1 & 0 & (f^{-1}-1)u_x \\ 0 & 1 & (f^{-1}-1)u_y \\ 0 & 0 & f^{-1} \end{bmatrix} \qquad (11)$$

with m_{ik} the points that one would have seen without change in focal length, $m_{f\,ik}$ the image points for some *relative* focal length f, $\mathbf{K}_f$ being the transformation between both.

The first thing to do is to retrieve the principal point u. Fortunately, this is easy for a camera with variable focal length, u being the only fixed point when varying the focal length without moving the camera. The affine camera calibration can for example be retrieved from two views with a different focal length and a pure translation between the two views, using the method described in [9].

From that point on we can use the *modulus constraint* to retrieve the focal length for supplementary views. This makes it possible to transform each image back to a normalized image (by canceling the change in focal length). From there one can use the method described in section 4 to get a full Euclidean calibration.

The *modulus constraint* is only valid for a normalized affine camera. By normalized it is meant that the change in focal length has been canceled. Stated differently the *modulus constraint* must be valid for a camera matrix $\mathbf{P}_{AkN} = \mathbf{K}_f^{-1}\mathbf{P}_{Ak}$. Writing down the characteristic equation we get an equation like Equation (8). Substituting the obtained coefficients in Equation (9) we obtain a 4^{th} order polynomial in f:

$$\alpha_4 f^4 + \alpha_3 f^3 + \alpha_2 f^2 + \alpha_1 f + \alpha_0 = 0 \qquad (12)$$

This gives 4 possible solutions. It can be proven that if f is a real solution, then $-f$ must also be a solution[1]. Imposing this to Equation (12) yields the following result:

$$f = \sqrt{\frac{\alpha_1}{\alpha_3}} \ . \qquad (13)$$

Now that we have retrieved f we can use K_f^{-1} to get normalized images and cameras. Then we can simply use the method described in section 4 to get a Euclidean calibration.

6 Experiments

In this section some results of self-calibration are given, both for the general motion method and the method allowing a variable focal length. For the first method experiments on synthetic data are presented. Results are compared with Hartley's [5]. The method is shown to work properly in the presence of realistic amounts of noise. For the second method results obtained from real data are shown. A 3D reconstruction of a scene is given which exhibits the Euclidean attributes of the real scenes (i.e. right angles, ...).

6.1 Experiment with general motion

For the method applicable with general motion – in particular affine reconstruction without relying on translation – two experiments will be presented here. One with a small number of views and one with a larger number of views.

[1] This is because the only constraint imposed is the *modulus constraint* (same modulus for all eigenvalues). A mirroring of the scene does not change the modulus of the eigenvalues, only the sign. Changing the sign of f has the same effect.

Noise	u_x	u_y	r_y^{-1}	$skew$	r_x^{-1}/r_y^{-1}
–	500.0	400.0	1000.0	-5.00	0.9000
0.0	494.9	402.1	1035.2	-16.11	0.8889
0.5	490.6	402.2	1058.3	-23.30	0.8812
1.0	475.0	403.4	1152.0	-59.65	0.8491
2.0	458.3	410.5	1180.3	-49.99	0.8372
4.0	438.6	396.4	1360.7	-97.95	0.7856

Table 1. Calibration results for 4 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 known values to be able to assess the quality of the calibration afterwards. Normally distributed 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 [5] with 15 views. To ease the comparison the same layout was used for the results.

The first experiment was carried out on 4 views[2]. 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 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.

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

[2]4 views being the minimum to retrieve the Euclidean calibration using this technique

Noise	u_x	u_y	r_y^{-1}	$skew$	r_x^{-1}/r_y^{-1}
0.0	500.0	400.0	1000.0	-5.00	0.9000
0.5	502.0	401.1	1000.9	-5.10	0.9000
1.0	499.3	398.3	997.9	-5.74	0.8992
2.0	501.6	397.8	978.0	1.37	0.9044
4.0	495.2	410.2	960.3	-8.24	0.8902

Table 2. Calibration results for 15 views

Figure 1. *The 3 images that were used to build a Euclidean reconstruction. The camera was translated between the first two views (the zoom was used to keep the size more or less constant). For the third image the camera was also rotated.*

the projective reconstruction.

6.2 Experiment with variable focal length

Here some results obtained from a real scene are presented. The scene consisted of two boxes and a cup. The images that were used can be seen in Figure 1. The scene was chosen to allow a good qualitative evaluation of the Euclidean reconstruction. The boxes have right angles and the cup is cylindrical. These characteristics must be preserved by a *Euclidean* reconstruction, but will in general not be preserved by an *affine* or *projective* reconstruction.

First a corner matcher was used to extract point correspondences between the three images. From these the method allowing varying focal lengths was used to obtain a Euclidean calibration of the cameras. Subsequently, an algorithm [11] was used to compute dense point correspondences. These were used to build the final 3D reconstruction using the previously recovered calibration.

Figure 2 shows two views of the reconstructed scene[3]. The left image is a front view, the right image a top view. Note especially from the top view, that 90^o angles are preserved and that the cup keeps its cylindrical form which is an indication of the quality of the *Euclidean* reconstruction.

[3]We preferred shading to the projection of the original texture on the model because this gives a better impression of the 3D structure.

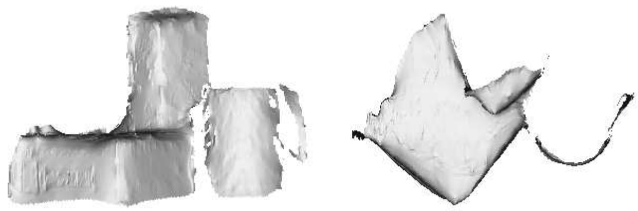

Figure 2. *front and top view of the reconstruction.*

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* [8]). From there on it is easy to get a Euclidean reconstruction. In contrast to the methods of Faugeras *et al* [3] and Hartley [5], 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 was briefly discussed and the feasibility was demonstrated by an experiment on real data. More details can be found in [10].

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 different methods could help. 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.

References

[1] M. Armstrong, A. Zisserman and P. Beardsley, Euclidean structure from uncalibrated images, *Proc. BMVC'94*.

[2] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, *Proc. ECCV'92*, pp.563-578.

[3] O. Faugeras, Q.-T. Luong and S. Maybank. Camera self-calibration: Theory and experiments, *Proc. ECCV'92*, pp.321-334.

[4] R. Hartley, Estimation of relative camera positions for uncalibrated cameras, *Proc. ECCV'92*, pp.579-587.

[5] R. Hartley, Euclidean reconstruction from uncalibrated views, Applications of invariance in Computer Vision, LNCS 825, Springer-Verlag, 1994.

[6] A. Heyden, K. Åström, Euclidean Reconstruction from Constant Intrinsic Parameters *Proc. ICPR'96.*

[7] M. Li, Camera Calibration of a Head-Eye System for Active Vision *Proc. ECCV'94*, pp.543–554.

[8] T. Moons, L. Van Gool, M. Van Diest and E. Pauwels, Affine reconstruction from perspective image pairs, Applications of Invariance in Computer Vision, LNCS 825, Springer-Verlag, 1994.

[9] M. Pollefeys, L. Van Gool and M. Proesmans, Euclidean 3D Reconstruction from Image Sequences with Variable Focal Lengths, *Proc. ECCV'96*, pp. 31-42.

[10] M. Pollefeys, L. Van Gool and André Oosterlinck, Self-calibration with the modulus constraint, Tech.Rep.KUL/ESAT/MI2/9609, 1996.

[11] M. Proesmans, L. Van Gool and A. Oosterlinck, Determination of optical flow and its discontinuities using non-linear diffusion, *Proc. ECCV'94*, pp. 295-304.

[12] C. Rothwell, G. Csurka and O.D. Faugeras, A comparison of projective reconstruction methods for pairs of views, *Proc. ICCV'95*, pp.932-937.

[13] M. Spetsakis and Y.Aloimonos, A Multi-frame Approach to Visual Motion Perception *International Journal of Computer Vision*, 6:3, 245-255, 1991.

[14] C. Zeller and O. Faugeras, Camera self-calibration from video sequences: the Kruppa equations revisited. Research Report 2793, INRIA, 1996.