Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters

Marc Pollefeys, Reinhard Koch and Luc Van Gool
ESAT-MI2, K.U.Leuven
Kardinaal Mercierlaan 94
B-3001 Heverlee, Belgium
firstname.lastname@esat.kuleuven.ac.be

Abstract

In this paper the feasibility of self-calibration in the presence of varying internal camera parameters is under investigation. A self-calibration method is presented which efficiently deals with all kinds of constraints on the internal camera parameters. Within this framework a practical method is proposed which can retrieve metric reconstruction from image sequences obtained with uncalibrated zooming/focusing cameras. The feasibility of the approach is illustrated on real and synthetic examples.

1 Introduction

In recent years, researchers have been studying self-calibration methods for cameras. Mostly completely unknown but constant intrinsic camera parameters were assumed. This has the disadvantage that to allow self-calibration the zoom can not be used and even focusing is prohibited. On the other hand, the proposed perspective model is often too general compared to the range of existing cameras. Mostly the image axes can be assumed orthogonal and often the aspect ratio is known (equal to one). Therefore a tradeoff can be made and by assuming these parameters to be known, one can allow (some of) the other parameters to vary throughout the image sequence.

Since it became clear that projective reconstructions could be obtained from image sequences alone [3, 5], researchers tried to find ways to upgrade these reconstructions to metric (i.e. Euclidean but unknown scale). Many methods where developed which assumed constant internal camera parameters [4, 15, 19, 7, 18, 13, 11, 12, 6]. Most of these methods are based on the **absolute conic** which is the only conic which stays fixed under all Euclidean transformations. This conic lays in the plane at infinity and its image is directly related to the internal camera parameters, hence the advantage for self-calibration [16].

Faugeras *et al* [4] (see also [15]) proposed to use the Kruppa equations which enforce that the planes through two camera centers which are tangent to the absolute conic should also be tangent to both of its images. Later on Zeller and Faugeras [19] proposed a more robust version of this method.

Heyden and Åström [7], Triggs [18] and Pollefeys and Van Gool [13] use explicit constraints which relate the absolute conic to its images. These formulations are especially interesting since they can easily be extended to deal with varying internal camera parameters.

Pollefeys *et al* [11, 12] also proposed a stratified approach which consists of first locating the plane at infinity using the modulus constraint (i.e. for constant internal camera parameters the infinity homography should be conjugated to a rotation matrix) and then calculating the absolute conic. Hartley [6] proposed another approach based on the minimization of the difference between the internal camera parameters for the different views.

So far not much work has been done on varying internal camera parameters. Pollefeys *et al* [10] also proposed a stratified approach for the case of a varying focal length, but this method required a pure translation as initialization. Recently Heyden and Åström [8] proved that self-calibration was possible when the aspect ratio was known and no skew was present. The self-calibration method proposed in their paper is of limited practical use since it is based on bundle adjustment which requires non-linear minimization over all reconstructed points and cameras simultaneously and the issue of initialization is not addressed.

In this paper a versatile self-calibration method is proposed which can deal with varying types of constraints. This will then be specialized towards the case were the focal length varies, possibly also the principal point. Section 2 of this paper first introduces some

basic principles. In Section 3 the actual method is developed. A simplified linear version is also given which can be used for initialization. Section 4 summarizes the complete procedure for metric reconstruction of arbitrarily shaped, rigid objects from an uncalibrated image sequence alone. The method is then validated through the experiments of Section 5. Section 6 concludes the paper and gives some directions for further research.

2 Some theory...

The projection of a scene onto an image can be modeled by the following equation:

$$\lambda m = \mathbf{P} M \tag{1}$$

where $m = [xy1]^\top$ is an image point and $M = [XYZ1]^\top$ is a scene point, $\mathbf{P}$ is the 3×4 camera projection matrix and λ is a scale factor. The camera projection matrix factorizes as follows:

$$\mathbf{P} = \mathbf{K}\,[\mathbf{R}|\text{-}\mathbf{R}\mathbf{t}] \text{ with } \mathbf{K} = \begin{bmatrix} f_x & s & u \\ & f_y & v \\ & & 1 \end{bmatrix} . \tag{2}$$

Here $(\mathbf{R}, \mathbf{t})$ denotes a rigid transformation (i.e. $\mathbf{R}$ is a rotation matrix and $\mathbf{t}$ is a translation vector), while the upper triangular calibration matrix $\mathbf{K}$ encodes the intrinsic parameters of the camera (i.e. f_x and f_y represent the focal length divided by the pixel width resp. height, (u, v) represents the principal point and s is a factor which is zero in the absence of skew).

It is a well-known result that from image correspondences alone the camera projection matrices and the reconstruction of the scene points can be retrieved up to a projective transformation [3, 5]. Note that without additional constraints nothing more can be achieved. This can be seen from the following equation. $\lambda m = \mathbf{P} M = (\mathbf{P}\mathbf{T}^{-1})(\mathbf{T} M) = \mathbf{P}' M'$ with $\mathbf{T}$ an arbitrary projective transformation. Therefore $(\mathbf{P}', M')$ is also a valid reconstruction from the image points m.

In general, however, some additional constraints are available. Some intrinsic parameters are known or can be assumed constant. This yields constraints which should be verified when $\mathbf{P}$ is factorized as in equation 2.

It can be shown that when no skew is present (i.e. $s = 0$ in equation 2), the ambiguity of the reconstruction can be restricted to metric. This can be derived from the following theorem:

Theorem

> The class of transformations which preserves the absence of skew is the group of similarity transformations.

A proof of this theorem can be found in [14]. Although this is theoretically sufficient, under practical circumstances often much more constraints are available and should be used.

3 Self-Calibration Methods

A practical way to obtain the calibration parameters from constraints on the internal camera parameters is through the absolute quadric [16] (introduced in computer vision by Triggs [18], see also [7, 13]). In space one quadric (of planes) exists which has the property to be invariant under rigid transformations. This quadric consists of planes tangent to the absolute conic. In a metric frame it is represented by a 4×4 symmetric rank 3 matrix $\Omega = \begin{bmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{bmatrix}$. If $\mathbf{T}$ transforms points $M \rightarrow \mathbf{T} M$ (and thus $\mathbf{P} \rightarrow \mathbf{P}\mathbf{T}^{-1}$), then it transforms $\Omega \rightarrow \mathbf{T}\Omega\mathbf{T}^\top$ (which can be verified to yield Ω when $\mathbf{T}$ is a similarity transformation). The projection of the absolute quadric in the image yields the dual image absolute conic:

$$\omega_k \simeq \mathbf{K}_k \mathbf{K}_k^\top \simeq \mathbf{P}_k \Omega \mathbf{P}_k^\top \tag{3}$$

independent of the chosen projective basis[1]. Therefore constraints on the internal camera parameters in $\mathbf{K}_i$ can be translated to constraints on the absolute quadric. If enough constraints are at hand only one quadric will satisfy them all, i.e. the *absolute quadric*. At that point the scene can be transformed to the metric frame (which brings Ω to its canonical form).

3.1 Non-linear Approach

Equation 3 can be used to obtain the metric calibration from the projective one. The dual image absolute conics ω_k should be parameterized in such a way that they enforce the constraints on the calibration parameters. For the absolute quadric $\mathbf{\Omega}$ a minimum parameterization (8 parameters) should be used. This can be done by putting $\Omega_{33} = 1$ and by calculating Ω_{44} from the rank 3 constraint.

An approximate solution to these equations can be obtained through non-linear least squares. The following criterion should be minimized:

$$\min \sum_{k=1}^{n} \left\| \mathbf{K}_k \mathbf{K}_k^\top - \mathbf{P}_k \Omega \mathbf{P}_k^\top \right\|_F^2 . \tag{4}$$

[1] Using Equation 2 this can be verified for a metric basis. Transforming $\mathbf{P} \rightarrow \mathbf{P}\mathbf{T}^{-1}$ and $\Omega \rightarrow \mathbf{T}\Omega\mathbf{T}^\top$ will not change the projection.

To obtain meaningful results $\mathbf{K}_k\mathbf{K}_k^\top$ and $\mathbf{P}_k\Omega\mathbf{P}_k^\top$ should both be normalized to have Frobenius norms equal to one.

To apply this self-calibration method to standard zooming/focusing cameras, some assumptions should be made. Often it can be assumed that there is no skew and that the aspect ratio is tuned to one. Often it can also be assumed that the principal point will be close to the center of the image. This leads to the following parameterizations for $\mathbf{K}_k$ (transform the images to have $(0,0)$ in the middle):

$$\mathbf{K}_k = \begin{bmatrix} f & 0 & u \\ & f & v \\ & & 1 \end{bmatrix} \text{ or } \mathbf{K}_k = \begin{bmatrix} f & 0 & 0 \\ & f & 0 \\ & & 1 \end{bmatrix} . \quad (5)$$

These parameterizations can be used in Equation 4. It will be seen in the experiments of Section 5 that this method gives good results on synthetic data as well as on real data.

3.2 Linear Approach

In the case were besides the skew ($s = 0$), both principal point and aspect ratio are (approximately) known a linear algorithm can be obtained by transforming the principal point $(u, v) \rightarrow (0, 0)$ and the aspect ratio $\frac{f_y}{f_x} \rightarrow 1$. If in addition one chooses $\mathbf{P}_1 = [\mathbf{I}|\mathbf{0}]$, Equation 3 can be rewritten as follows:

$$\lambda_k \begin{bmatrix} f_k^2 & 0 & 0 \\ 0 & f_k^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{P}_k \begin{bmatrix} f_1^2 & 0 & 0 & a_1 \\ 0 & f_1^2 & 0 & a_2 \\ 0 & 0 & 1 & a_3 \\ a_1 & a_2 & a_3 & \|\mathbf{a}\|^2 \end{bmatrix} \mathbf{P}_k^\top \quad (6)$$

From the left-hand side of Equation 6 it can be seen that the following equations have to be satisfied: $\omega_{11} = \omega_{22}, \omega_{12} = \omega_{13} = \omega_{23} = 0$. This can thus be imposed on the right-hand side, yielding $4(n-1)$ linear equations in f_1^2, a_1, a_2, a_3 and $\|\mathbf{a}\|^2$. The rank 3 constraint can be imposed by taking the closest rank 3 approximation (using SVD for example).

When only two views are available the solution is only determined up to a one parameter family of solutions. Imposing the rank 3 constraint in this case results in up to 4 possible solutions.

4 Overview of the Reconstruction Algorithm

The proposed self-calibration method is embedded in a system to automatically model metric reconstructions of rigid 3D objects from uncalibrated video sequences. The complete procedure for metric 3D reconstruction is summarized here.

1. Corners are matched between the different images of the sequence and the projective camera matrices $\mathbf{P}_k$ are robustly estimated from them. A more detailed explanation of this approach can be found in [1].

2. Self-Calibration

   (a) First the focal lengths f_k are estimated using the linear algorithm (assuming that there is no skew and that the principal point is in the middle of the image). These results are refined using the nonlinear algorithm .

   (b) (optional) Not only f_k, but also the principal points (u_k, v_k) can be estimated for all the views.

   (c) Upgrade from projective $\mathbf{P}_k$ to metric using the obtained absolute quadric.

3. The epipolar geometry is used to obtain a pairwise image rectification where all epipolar lines lay along the image scan lines. Dense correspondence matches are then computed along each scan line using hierarchical block-matching and dynamic programming [2].

4. Reconstruction of a dense 3D model based on metric $\mathbf{P}_k$ and dense correspondence maps. The dense image matches are triangulated in space using the now calibrated cameras and a textured 3D surface model is obtained that can easily be loaded into computer graphic systems for realistic rendering [9].

5 Experiments

In this section a number of experiments are described. First some synthetic image sequences were used to assess the quality of the algorithm under simulated circumstances. Then results are given for two outdoor video sequences. Both sequences were taken with a standard semi-professional camcorder that was moved freely around the objects. Sequence 1 was filmed with constant camera parameters –like most algorithms require. The new algorithm –which doesn't impose this– could therefore be tested on this. A second sequence was recorded with varying internal camera parameters. A zoom factor ($2\times$) was applied while filming.

5.1 Simulations

Extensive synthetic experiments have been carried out to analyze the noise sensitivity of the method using different constraints. From these we could conclude that for small amounts of noise the more complex

Figure 1: First and last image of Sequence 1

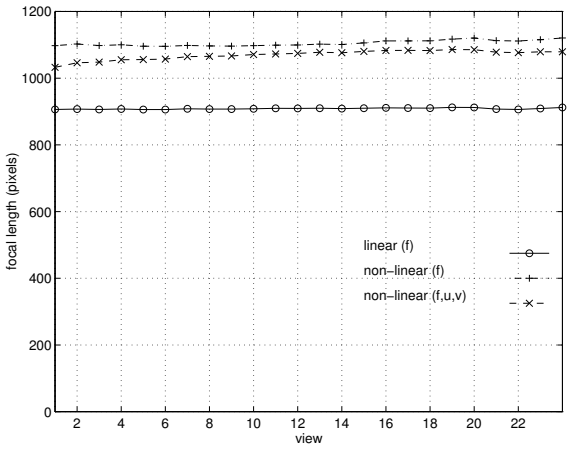

Figure 2: focal length (in pixels) versus views for the different algorithms

models should be preferred. If more noise is added, the simple model gives the best results. This is due to the low redundancy of the system of equations for the models which, beside the focal length, also try to estimate the position of the principal point.

Another experiment was carried out to evaluate the performance of the algorithm for different sequence length. The longer the sequence the better the results, especially when few constraints are used. More details can be found in [14].

5.2 Outdoor Sequence 1

The first sequence showing part of an old castle was filmed with a fixed zoom/focus. It is therefore a good test for the algorithms presented in this paper to check if they indeed return constant intrinsic parameters for this sequence. In Figure 1 the first and the last of the

	angle ($\pm$std.dev.)
parallelism	1.0 ± 0.6 degrees
orthogonality	92.5 ± 0.4 degrees

Table 1: Results of metric measurements on the reconstruction

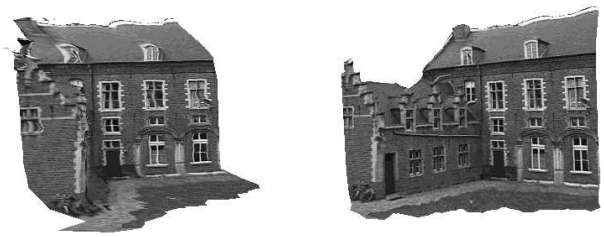

Figure 3: Perspective views of reconstruction

24 images of the sequence are shown.

In Figure 2 the focal length for every view is plotted for the different algorithms. The calculated focal lengths are almost constant as it should be. When the principal point is not fixed by the algorithm, it varies over a few tens of pixels. In this case it seems that the projective calibration was not accurate enough to allow the retrieval of the principal point and it is better to choose the simplified algorithm.

To judge the visual quality of the reconstruction, different perspective views of the model were computed and displayed in Figure 3. The result is rather convincing, although at the moment only a partial model from a single image pair is reconstructed.

A quantitative assessment of the metric properties can be made by explicitly measuring angles directly on the object surface. For this experiment lines were placed aligned with the windows. Therefore these lines should be either parallel (angle between them should be 0 degree) or perpendicular to each other (angle between them should be 90 degree). The measurement on the object surface shows that this is indeed close to the expected values (see Table 2).

5.3 Outdoor Sequence 2

This sequence shows a stone pillar with curved surfaces. While filming and moving away the zoom was changed to keep the image size of the object constant. The focal length was not changed between the two first images, then it was changed more or less linearly. From the second image to the last image the focal

Figure 4: First and last image of sequence 2

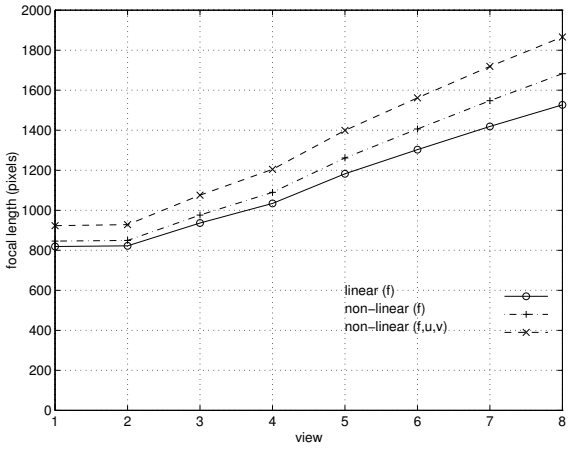

Figure 5: focal length (in pixels) versus views for the different algorithms

length has been doubled (if the markings on the camera can be trusted). The first and the last image of the sequence (8 images) can be seen in Figure 4. Notice that the perspective distortion is much more visible in the first image (wide angle) than the last image (longer focal length).

In Figure 5 the focal length for every view is plotted for the different algorithms. It can be seen that the calculated values of the focal length correspond to what could be expected. When not fixed the principal point moves substantially (up to 100 pixels). It is therefore probable that too much noise is present to allow us to estimate the principal point accurately.

In Figure 6 a perspective view of the reconstruction is given, rendered both shaded and with surface texture mapping. The shaded view shows that even most of the small details of the object are retrieved. Figure 7 shows a left and a right side view of the reconstructed object. Although there is some distortion at the outer boundary of the object, a highly realistic impression of the object is given. Note the arbitrarily shaped free-form surface that has been reconstructed. A quantitative assessment of the metric properties for the pillar is not so easy because of the curved surfaces.

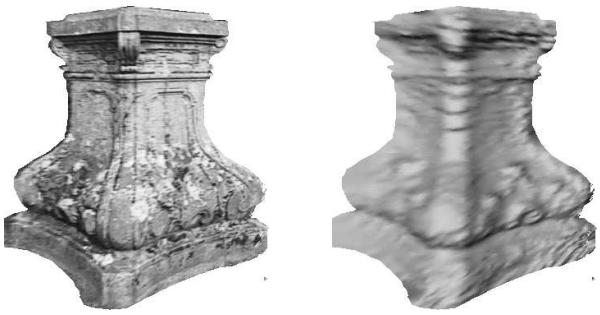

Figure 6: Perspective view of the reconstruction (with texture and with shading).

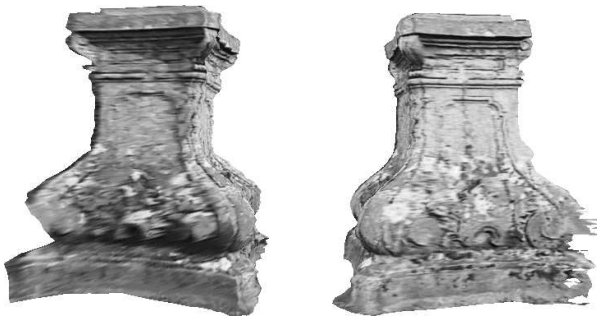

Figure 7: Left and right perspective view of the reconstruction.

It is, however, possible to measure some distances on the real object as reference lengths and compare them with the reconstructed model. In this case it is possible to obtain a measure for the absolute scale and verify the consistency of the reconstructed length within the model. For this comparison a network of reference lines was placed on the original object and 27 manually measured object distances were compared with the reconstructed distances on the model surface. Due to the increased reconstruction uncertainty at the outer object silhouette some distances show a larger error than the interior points. Averaging all 27 measured distances gave a consistent scale factor of 40.25 with a standard deviation of 5.4% overall. For the interior distances, the reconstruction error dropped to 2.3%. These results demonstrate the metric quality of the reconstruction even for complicated surface shapes and varying focal length.

6 Conclusions

This paper focussed on self-calibration and metric reconstruction in the presence of varying and unknown internal camera parameters. The calibration models used in previous research are on one hand too restrictive in real imaging situations (constant parameters)

and on the other hand too general (all parameters unknown). The more pragmatic approach which is followed in this paper results in more flexibility.

A versatile self-calibration method which can work with different types of constraints (some of the internal camera parameters constant or known) was derived. This method was then specialized towards the practically important case of a zooming/focusing camera (without skew and an aspect ratio $\frac{f_y}{f_x} = 1$). Both known and unknown principal point were considered. It is proposed to always start with the principal point in the center of the image and use the linear algorithm. The non-linear minimization is then used to refine the results, possibly –for longer sequences– allowing the principal point to be different for each image. This can however degrade the results if the projective calibration was not accurate enough, the sequence not long enough, or the motion sequence critical towards the set of constraints.

Acknowledgements

We would like to thank Andrew Zisserman and his group from Oxford for supplying us with robust projective reconstruction software. A specialization grant from the Flemish Institute for Scientific Research in Industry (IWT) and the financial support from the EU ACTS project AC074 'VANGUARD' are also gratefully acknowledged.

References

[1] P. Beardsley, P. Torr and A. Zisserman 3D Model Acquisition from Extended Image Sequences *Proc.ECCV'96.*

[2] L. Falkenhagen, Depth Estimation from Stereoscopic Image Pairs assuming Piecewise Continuous Surfaces, *Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Productions*, Hamburg, Germany, 1994.

[3] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig, *Proc.ECCV'92.*

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

[5] R. Hartley, Estimation of relative camera positions for uncalibrated cameras, *Proc.ECCV'92.*

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

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

[8] A. Heyden, K. Åström, Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point, *Proc.CVPR'97.*

[9] R. Koch, 3-D Surface Reconstruction from Stereoscopic Image Sequences, *Proc.ICCV'95.*

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

[11] M. Pollefeys, L. Van Gool and A. Oosterlinck, The Modulus Constraint: A New Constraint for Self-Calibration, *Proc.ICPR'96.*

[12] M. Pollefeys and L. Van Gool, A stratified approach to self-calibration, *Proc.CVPR'97.*

[13] M. Pollefeys and L. Van Gool, Self-calibration from the absolute conic on the plane at infinity, *Proc.CAIP'97.*

[14] M. Pollefeys and L. Van Gool, Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Paramaters, TR9707, ESAT-MI2, K.U.Leuven, 1997.

[15] Q.-T. Luong and O. Faugeras, Self Calibration of a moving camera from point correspondences and fundamental matrices, IJCV, vol.22-3, 1997.

[16] J.G. Semple and G.T. Kneebone, Algebraic Projective Geometry, Oxford University Press, 1952.

[17] P. Sturm, Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction, *Proc.CVPR'97.*

[18] B. Triggs, The Absolute Quadric, *Proc.CVPR'97.*

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