Self-calibration of zooming cameras observing an unknown planar structure

Ezio Malis and Roberto Cipolla
University of Cambridge, Engineering Department
Trumpington Street, CB2 1PZ Cambridge, United Kingdom
em240,cipolla@eng.cam.ac.uk

Abstract

In this paper, we propose a new self-calibration technique for cameras with changing zoom observing only a planar structure. The method does not need any metric or topologic knowledge about the structure since it is based on the estimation of the collineations existing between several views of a plane (thus only image correspondences are needed). The constraints existing between all the collineations are imposed using a very simple and efficient technique which does not need the solution of a complex optimisation problem. Finally, even if the structure of the plane is unknown it must be the same for all the images and this provides some constraints which allow the recovering of the varying focal length.

1. Introduction

Camera self-calibration from views of a generic scene has been widely investigated and the two main approaches are based on the properties of absolute conics [14] [11] or on some algebraic error [8] [4]. Depending on the a priori information provided the self-calibration algorithms can be classified as follows. Algorithms that use some knowledge of the observed scene: identifiable targets of known shape [9], metric structure of planes [12]. Algorithms that exploit particular camera motions: translating camera or rotating camera [5]. Algorithms that have some knowledge on the camera parameters: some fixed camera parameters (i.e. skew zero, unit ratio etc.), varying camera parameters [11] [10] [1]. In this paper, we propose a self-calibration technique for zooming cameras observing only an unknown planar structure. The particular geometry of features lying on planes is often the reason for the inaccuracy of many computer vision applications (structure from motion, self-calibration) if it is not taken explicitly into account in the algorithms. Introducing some knowledge about the coplanarity of the features and about their structure (metric or topological) can improve the quality of the estimates [13]. However, the only prior geometric knowledge on the features that will be used here is their coplanarity. Two views of a plane are related by a collineation. Using multiple views of a plane we obtain a set of collineations which are not independent. In order to avoid solving non-linear optimisation problems, the constraints existing within a set of collineation and between sets have often been neglected. However, these multi-view constraints can be used to improve the estimation of the collineations matrices as in [16], where multiple planes are supposed to be viewed in the images. In this papers the constraints are imposed using a very simple and efficient technique which does not need the solution of a complex optimisation problem. Imposing the constraint is useful since it allows the reduction of the geometric error in the reprojected features and provides a consistent set of collineations which can be used for camera self-calibration. Camera self-calibration from planar scenes with known metric structure has been investigated [12]. However, it is interesting to develop flexible techniques which do not need any a priori knowledge about the camera motion as in [5] or metric knowledge of the planar scene. Methods for self-calibrating a camera from views of planar scenes without knowing their metric structure were proposed in [15] and [7]. In [7] the internal parameters of the camera are supposed to remain constant. In this paper we investigate how to improve the self-calibration from planar scenes of a camera with varying focal length.

2. The model of the camera

A camera performs a perspective projection of a point $\mathbf{x} \in \mathbb{P}^3$ to an image point $\mathbf{p} \in \mathbb{P}^2$ measured in pixels: $\mathbf{p} \propto \mathbf{K} \begin{bmatrix} \mathbf{R} & \mathbf{t} \end{bmatrix} \mathbf{x}$, where $\mathbf{R}$ and $\mathbf{t}$ represent the displacement between the frame $\mathcal{F}$ attached to the camera and an absolute coordinate frame $\mathcal{F}_0$, and $\mathbf{K}$ is a (3×3) matrix containing the intrinsic parameters of the camera:

$$\mathbf{K} = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where u_0 and v_0 are the coordinates of principal point (in pixels), f is the focal length (in meters), k_u and k_v are the magnifications respectively in the $\vec{u}$ and $\vec{v}$ direction (in pixels/meters). In this paper we will suppose also that the skew is zero which is in general a good approximation.

3. Two-view geometry of a plane

Two views of a plane are related by a collineation matrix. Indeed, the image coordinates $\mathbf{p}_{ik}$ of the point $\mathcal{P}_k$ in the image $\mathcal{I}_i$ can be obtained from the image coordinates $\mathbf{p}_{jk}$ of the point $\mathcal{P}_k$ in the image $\mathcal{I}_j$:

$$\mathbf{p}_{ik} \propto \mathbf{G}_{ij}\mathbf{p}_{jk} \tag{2}$$

The collineation matrix is a (3×3) matrix defined up to scalar factor which can be written as:

$$\mathbf{G}_{ij} \propto \mathbf{K}_i\mathbf{H}_{ij}\mathbf{K}_j^{-1} \tag{3}$$

where $\mathbf{H}_{ij}$ is the corresponding homography matrix in the Euclidean space. Homography and collineation are generally used to indicate the same projective transformation from $\mathbb{P}^n$ to $\mathbb{P}^n$ (in our case $n = 2$). In this paper we will use the term “homography” to indicate a collineation expressed in the Euclidean space. The homography matrix can be written as a function of the camera displacement [2]:

$$\mathbf{H}_{ij} = \mathbf{R}_{ij} + \frac{\mathbf{t}_{ij}\ \mathbf{n}_j^T}{d_j} \tag{4}$$

where $\mathbf{R}_{ij}$ and $\mathbf{t}_{ij}$ are respectively the rotation and the translation between the frames $\mathcal{F}_i$ and $\mathcal{F}_j$, $\mathbf{n}_j$ is the normal to the plane π expressed in the frame $\mathcal{F}_j$ and d_j is the distance of the plane π from the origin of the frame $\mathcal{F}_j$. From equation (3) the $\mathbf{H}_{ij}$ can be estimated from $\mathbf{G}_{ij}$ knowing the camera internal parameters of the two cameras:

$$\mathbf{H}_{ij} \propto \mathbf{K}_i^{-1}\mathbf{G}_{ij}\mathbf{K}_j \tag{5}$$

However, it should be noticed that the Euclidean homography matrix is *not* defined up to a scale factor since its median singular value must be equal to 1 (see [7] for details).

From equation (4) it is easy to verify that the homography matrix satisfies the constraint $\forall k > 0$ (where $[\mathbf{n}_i]_\times$ is the skew symmetric matrix associated with vector $\mathbf{n}_i$):

$$[\mathbf{n}_i]_\times^k\ \mathbf{H}_{ji}^T = \mathbf{H}_{ij}\ [\mathbf{n}_j]_\times^k \tag{6}$$

If $k = 1$, the matrix $[\mathbf{n}_i]_\times \mathbf{H}_{ji}^T$ has similar properties to the essential matrix (i.e. $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$). Indeed, this matrix has two equal singular values and one equal to zero. This means each homography places two constraints on the internal camera parameters [6] which can be used for self-calibration as in [10]. Another very important relation is the following:

$$[\mathbf{n}_i]_\times = \mathbf{H}_{ij}\ [\mathbf{n}_j]_\times\ \mathbf{H}_{ij}^T \tag{7}$$

Indeed, since $\det(\mathbf{M})\mathbf{M}\,[\mathbf{v}]_\times\,\mathbf{M}^T = \left[\mathbf{M}^{-T}\mathbf{v}\right]_\times$ it means that the normals to the plane are related by:

$$\mathbf{n}_i = \mathbf{Q}_{ij}\mathbf{n}_j \tag{8}$$

where:

$$\mathbf{Q}_{ij} = \det(\mathbf{H}_{ij})\ \mathbf{H}_{ij}^{-T} \tag{9}$$

These important equations will be extended to the multi-view geometry in the next section

4 Multi-view geometry

4.1 The super-collineation matrix

If m images of an unknown planar structure are available, it is possible to compute m^2 collineations. Let us define the super-collineation matrix as follows:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{11} & \cdots & \mathbf{G}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{G}_{m1} & \cdots & \mathbf{G}_{mm} \end{bmatrix} \tag{10}$$

with $\dim(\mathbf{G}) = (3m, 3m)$ and $\text{rank}(\mathbf{G}) = 3$. The rank can not be less than three since $\mathbf{G}_{ii} = \mathbf{I}_3\ i \in \{1, ..., m\}$, and cannot be more than three since each row of the matrix can be obtained from a linear combination of three others rows:

$$\mathbf{G}_{ij} = \mathbf{G}_{ik}\mathbf{G}_{kj} \qquad \forall i, j, k \in \{1, 2, 3, ..., m\} \tag{11}$$

The constraints (11) can be summarised by the constraint:

$$\mathbf{G}^2 = m\ \mathbf{G} \tag{12}$$

Then, matrix $\mathbf{G}$ has 3 nonzero equal eigenvalues $\lambda = m$ and $3(m-1)$ null eigenvalues. If we can impose the constraint $\mathbf{G}^2 = m\ \mathbf{G}$ (with $\mathbf{G}_{ii} = \mathbf{I}_3\ i = 1, ..., m$) then it is equivalent to imposing the constraints $\mathbf{G}_{ij} = \mathbf{G}_{ik}\mathbf{G}_{kj}$.

In order to impose the constraint (12), we use the algorithm proposed in [7] which treats all the images with the same priority without using any key image and forces the rank 3 constraint on $\mathbf{G}$.

4.2 The super-homography matrix

Let us define the super-homography matrix in the Euclidean space as:

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1m} \\ \vdots & \ddots & \vdots \\ \mathbf{H}_{m1} & \cdots & \mathbf{H}_{mm} \end{bmatrix} \tag{13}$$

with $\dim(\mathbf{H}) = (3m, 3m)$ and $\text{rank}(\mathbf{H}) = 3$. The super-homography matrix can be obtained from the super-collineation matrix knowing the parameters of the cameras:

$$\mathbf{H} = \mathbf{K}^{-1}\mathbf{G}\mathbf{K} \tag{14}$$

where ($\dim(\mathbf{K}) = (3m, 3m)$ and $\mathrm{rank}(\mathbf{K}) = 3m$):

$$\mathbf{K} = \begin{bmatrix} \mathbf{K}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{K}_m \end{bmatrix} \quad (15)$$

is the matrix containing the internal parameters of the cameras. If the constraint $\mathbf{G}^2 = m\mathbf{G}$ is imposed, then the constraint $\mathbf{H}^2 = m\mathbf{H}$ is automatically imposed which means that the following constraints are satisfied:

$$\mathbf{H}_{ij} = \mathbf{H}_{ik}\mathbf{H}_{kj} \qquad \forall i, j, k \in \{1, 2, 3, ..., m\} \quad (16)$$

The supery-homography matrix is normalised by setting the median singular value of each homography to one. After normalisation, the homography matrix is decomposed as:

$$\mathbf{H} = \mathbf{R} + \mathbf{T}\mathbf{N}^T \quad (17)$$

with $\dim(\mathbf{R}) = (3m, 3m)$, $\mathrm{rank}(\mathbf{R}) = 3$, $\dim(\mathbf{T}) = (3m, m)$ and $\mathrm{rank}(\mathbf{T}) = m$, $\dim(\mathbf{N}) = (3m, m)$ and $\mathrm{rank}(\mathbf{N}) = m$. Matrix $\mathbf{R}$ is a symmetric matrix, $\mathbf{R} = \mathbf{R}^T$ and $\mathbf{R}^2 = m\mathbf{R}$. In [2] is presented a method for decomposing the homography matrix, computed from two views of a planar structure, following equation (4). In general, there are two possible solutions but the ambiguity can be solved by adding more images. In [7] we presented a method to decompose any set of homography matrices.

4.3 Camera self-calibration

In this section we use the properties of the set of homography matrices to self-calibrate the focal length of the camera. Each independent homography will provide us two constraints on the parameters. However two constraints are fixed by the normals to the plane. Therefore, the total number of constraints which can be obtained from m images is $2(m-1)$. Indeed, if σ_{ij}^{I} and σ_{ij}^{II} are the two non-zero singular values of $[\mathbf{n}_i]_\times \mathbf{H}_{ji}^T$ our self-calibration method is based on the minimisation of the following cost function [4][10]:

$$\mathcal{C} = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\sigma_{ij}^{I} - \sigma_{ij}^{II}}{\sigma_{ij}^{I}} \quad (18)$$

Using this cost function we need at least:

- 3 independent homography matrices (4 images) to recover the 4 different focal lengths (supposing the ratio and the principal point approximatively known);
- 4 independent homography matrices (5 images) to recover the 5 different focal lengths and the fixed ratio (with the principal point approximatively known);
- 6 independent homography matrices (7 images) to recover the 7 different focal lengths, the ratio and the the principal point.

5. Experimental results

5.1. Self-calibration of a camera without zooming

In this experiment we took a sequence (10 images) of a calibration grid (in order to have a ground truth) using a camera with 7mm focal length. In our sequence the focal length did not vary but we will suppose it unknown for each image as if the camera was zooming. Figure 1 shows three images of the sequence.

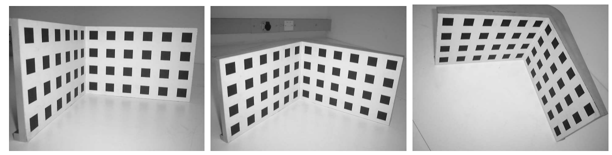

Figure 1. Images of the sequence taken with a camera with fixed focal length

The camera was calibrated (in order to have a ground truth) with the standard Faugeras-Toscani method [3] and the obtained focal length was $f = 685$. In order to test our self-calibration technique, the ratio k_u/k_v is fixed to one and the principal point is supposed to be in the center of the image. Thus, the unknowns are the 10 focal lengths. The results obtained using the left plane of the calibration grid (similar results have been obtained using the right plane) are summarised in Table 1. The starting focal length was $f_0 = 1000$ for all the unknown f_i.

f	685	error
f_1	685.7	+0.11 %
f_2	674.7	-1.49 %
f_3	692.1	+1.04 %
f_4	677.1	-1.15 %
f_5	697.3	+1.80 %
f_6	680.9	-0.60 %
f_7	684.4	-0.09 %
f_8	688.0	+0.44 %
f_9	675.0	-1.46 %
f_{10}	672.4	-1.843 %

Table 1. Self-calibration of the focal lengths using a camera without zooming

The results are very good since the maximal error is 1.8% of the focal length measured with the standard Faugeras-Toscani method. The mean of all the focal lengths is 682.8 (only 0.3% of f) and the standard deviation is 8.2 (only 1.2% of f). Even if in this experiment the focal length did not vary it was recovered from a starting focal length of 1000 pixels (which means an initial error of 45%).

5.2. Self-calibration of a zooming camera

In this experiment we took a sequence (10 images) of a calibration grid using a zooming camera. Figure 2 shows three images of the sequence.

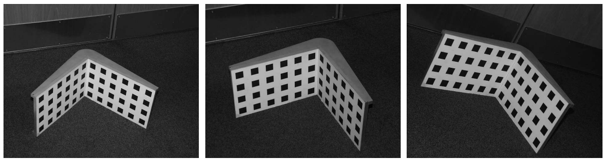

Figure 2. Images of the sequence taken with a zooming camera

The results obtained using the left plane of the calibration grid (similar results have been obtained using the right plane) are summarised in Table 2. The starting focal length was again $f_0 = 1000$ for all the unknown f_i.

f	Faugeras-Toscani	proposed method	error
f_1	1407.3	1491.0	-5.95 %
f_2	1835.0	1950.1	-6.27 %
f_3	1195.2	1211.9	-1.39 %
f_4	1491.6	1471.8	1.32 %
f_5	1337.0	1393.0	-4.18 %
f_6	1158.0	1233.1	-6.48 %
f_7	985.3	1012.2	-2.72 %
f_8	1534.1	1608.9	-4.87 %
f_9	1844.9	1929.1	-4.56 %
f_{10}	1839.0	1904.8	-3.57 %

Table 2. Self-calibration of the focal lengths with a zooming camera

Considering that in our self-calibration algorithm the principal point was supposed to be in the center of the image, the results are satisfactory and the 3D reconstruction of the grid can be done with sufficient accuracy.

6. Conclusions

In this paper we presented a new technique to self-calibrate cameras with varying focal length. Our method does not need any a priori knowledge of the metric structure of the plane. Moreover, we impose the constraints existing within a set of collineation matrices computed from multiple views of an unknown planar structure obtaining a consistent set of collineations. The method was tested using real images with a ground truth and the obtained results are very good. The method could be easily improved by using an error model.

7 Acknowledgements

This work was supported by an EC (ESPRIT) grant no. LTR26247 (VIGOR).

References

[1] L. de Agapito, R. Hartley, and E. Hayman. Linear self-calibration of a rotating and zooming camera. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 15–21, 1999.

[2] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *Int. Jour. of Pattern Recognition and Artificial Intelligence*, 2(3):485–508, 1988.

[3] O. Faugeras and G. Toscani. The calibration problem for stereo. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 15–20, June 1986.

[4] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In G. Sandini, editor, *Proc. European Conf. on Computer Vision*, volume 588 of *Lecture Notes in Computer Science*, pages 579–587. Springer-Verlag, May 1992.

[5] R. Hartley. Self-calibration from multiple views with a rotating camera. In *Proc. European Conf. on Computer Vision*, pages 471–478, May 1994.

[6] R. Hartley. Minimising algebraic error in geometric estimation problem. In *Proc. IEEE Int. Conf. on Computer Vision*, pages 469–476, 1998.

[7] E. Malis and R. Cipolla. Multi-view constraints between collineations: application to self-calibration from unknown planar structures. In *European Conference on Computer Vision*, June 2000.

[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] J. Mendelsohn and K. Daniilidis. Constrained self-calibration. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 581–587, 1999.

[10] P. Mendonca and R. Cipolla. A simple techinique for self-calibration. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 500–505, 1999.

[11] M. Pollefeys, R. Koch, and L. VanGool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, 32(1):7–25, August 1999.

[12] P. Sturm and S. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 432–437, 1999.

[13] R. Szeliski and P. Torr. Geometrically constrained structure from motion: Points on planes. In *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments (SMILE)*, pages 171–186, Germany, June 1998.

[14] B. Triggs. Autocalibration and the absolute quadric. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 609–614, 1997.

[15] B. Triggs. Autocalibration from planar scenes. In *Proc. European Conf. on Computer Vision*, pages 89–105, 1998.

[16] L. Zelnik-Manor and M. Irani. Multi-view subspace constraints on homographies. In *Proc. IEEE Int. Conf. on Computer Vision*, volume 1, pages 710–715, September 1999.