Self-Calibration of a Rotating Camera with Varying Intrinsic Parameters

L. de Agapito, E. Hayman and I. Reid
Department of Engineering Science, Oxford University
Parks Road, Oxford, OX1 3PJ, UK
[lourdes|hayman|ian]@robots.ox.ac.uk

Abstract

We present a method for self-calibration of a camera which is free to rotate and change its intrinsic parameters, but which cannot translate. The method is based on the so-called infinite homography constraint which leads to a non-linear minimisation routine to find the unknown camera intrinsics over an extended sequence of images. We give experimental results using real image sequences for which ground truth data was available.

1 Introduction

Camera calibration has always been the subject of research in the field of machine vision, however it was only relatively recently that the possibility of *self-calibration* of a camera simply by observing an unknown scene was realised and explored. The first major work to consider the problem was [3], which showed that self-calibration was theoretically and practically feasible for a camera moving through an unknown scene with constant but unknown intrinsics. Since that time various methods have been developed to deal with different situations. Table 1 summarises the major contributors to date.

In this paper we address one of the few cases which has not yet been explored, that of a stationary camera which may rotate and change its intrinsics. This lack of attention is somewhat surprising since this situation is one which occurs frequently in a variety of circumstances: surveillance devices and cameras used for broadcasts of (for example) sporting events are almost invariably fixed in location but free to rotate and zoom, and hand-held camcorders are very often panned from a single viewpoint. Note that although we address the case where the camera undergoes pure rotation (i.e. about its optic centre), in practice the method is applicable whenever the rotation arm is very small relative to the distance of the scene.

Our work is most closely related to the works of Hartley [4] and Pollefeys et al. [10], but differs from the former in that we consider the case of varying rather than fixed intrinsics, and from the latter in that we consider pure rotations, a case not handled by that work.

The paper is organised as follows. We begin with a description of the camera model (section 2), then derive a constraint on the dual of the image of the absolute conic, which forms the basis of our approach (section 3). We relate our work to [15, 10] in section 3.1. Readers familiar with the theory in [4, 7, 15, 10] may well be able to skip sections 2 – 3 and move directly to sections 4 and 5 which respectively describe the algorithm for self-calibration and experiments on real imagery.

	Constant intrinsics	Varying intrinsics
Known rotation	McLauchlan and Murray [9], Stein [14], Du and Brady [2]	—
Unknown rotation	Hartley [4], Zisserman *et al.* [16]	*
Unknown general motion	Maybank, Faugeras *et al.* [3, 8], Triggs [15], Pollefeys *et al.* [11], Heyden and Aström [5]	Heyden and Aström [6], Pollefeys *et al.* [10]

Table 1: A summary of some of the best work done on camera calibration from unknown scenes. The (*) indicates our contribution.

2 Camera model

The projection of scene points onto an image by a perspective camera may be modelled by the equation $\mathbf{x} = \mathtt{P}\mathbf{X}$, where $\mathbf{x} = [x\ y\ w]^\top$ are the image points in homogeneous coordinates, $\mathbf{X} = [X\ Y\ Z\ 1]^\top$ are the world points and $\mathtt{P}$ is the 3×4 camera projection matrix. The matrix $\mathtt{P}$ is a rank-3 matrix which may be decomposed as $\mathtt{P} = \mathtt{K}[\mathtt{R}| - \mathtt{R}\mathbf{t}]$, where the rotation $\mathtt{R}$ and the translation $\mathbf{t}$ represent the Euclidean transformation between the camera and the world coordinate systems and the matrix $\mathtt{K}$ is an upper triangular matrix which encodes the internal parameters of the camera in the form

$$\mathtt{K} = \begin{bmatrix} \alpha_u & k & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} . \quad (1)$$

The elements α_u and α_v represent the focal length of the camera expressed in horizontal and vertical pixel units respectively. The aspect ratio is $r = \alpha_v/\alpha_u$. The principal point is (u_0, v_0) and k is a skew parameter which is a function of the angle between the horizontal and vertical axes of the sensor array.

In this paper we will address the problem of computing the calibration matrix of a stationary camera undergoing pure rotation with vaying internal parameters. If we choose the origin of the coordinate system to be the optic centre of the camera, common to all views, we may write the projection matrices in the form

$$\mathtt{P}_i = \mathtt{K}_i [\mathtt{R}_i|0] , \quad (2)$$

where $\mathtt{K}_i$ is the calibration matrix for each view and $\mathtt{R}_i$ describes the orientation of the camera with respect to the chosen reference frame. Under such circumstances, a world point $\mathbf{X} = [X\ Y\ Z\ 1]^\top$ is mapped onto the image point $\mathbf{x} = \mathtt{K}_i [\mathtt{R}_i|0] [X\ Y\ Z\ 1]^\top =$

$\mathtt{K}_i\mathtt{R}_i\,[X\ Y\ Z]^\top$. Note that since the fourth column of the projection matrix is always zero, the last coordinate of the world points $\mathbf{X}$ is irrelevant, so we write $\overline{\mathbf{X}} = [X\ Y\ Z]^\top$, and the mapping of world to image points may be conveniently expressed by the 3×3 projective transformation $\overline{\mathtt{P}}_i = \mathtt{K}_i\mathtt{R}_i$.

3 Rotation of a stationary camera

In this section we prove an extension to Hartley's so-called *infinite homography constraint* [4], in which we consider a camera with changing intrinsics. In particular we show that the dual of the image of the absolute conic ω^* is related between views i and j by the following equation:

$$\boldsymbol{\omega}_j^* = \mathtt{H}_{ij}\boldsymbol{\omega}_i^*\mathtt{H}_{ij}^\top, \tag{3}$$

where $\mathtt{H}_{ij}$ is the homography that maps corresponding points from view i to j. This result can also be found in [7], but was not used for the same purpose as in our work.

To begin the proof, let the world coordinate system be aligned with the camera in the first frame. We may then write the projection matrices for the different views as

$$\overline{\mathtt{P}}_0 = \mathtt{K}_0 \qquad \overline{\mathtt{P}}_i = \mathtt{K}_i\mathtt{R}_i. \tag{4}$$

Given a 3D world point $\overline{\mathbf{X}}$, its projections onto two different images will be $\mathbf{x}_i = \mathtt{K}_i\mathtt{R}_i\overline{\mathbf{X}}$ and $\mathbf{x}_j = \mathtt{K}_j\mathtt{R}_j\overline{\mathbf{X}}$, so eliminating $\overline{\mathbf{X}}$ yields the transformation relating corresponding points

$$\mathbf{x}_j = \mathtt{K}_j\mathtt{R}_j\mathtt{R}_i^{-1}\mathtt{K}_i^{-1}\mathbf{x}_i \tag{5}$$

Therefore, in the case of a stationary camera there exists a 2D homography which maps corresponding points in two views:

$$\mathtt{H}_{ij} = \mathtt{K}_j\mathtt{R}_{ij}\mathtt{K}_i^{-1}. \tag{6}$$

$\mathtt{H}_{ij}$ is in fact the infinite homography $\mathtt{H}_\infty$ between views i and j, i.e. the point homography between image planes induced by the plane at infinity. Therefore, in the case of a rotating camera the infinite homography is an observable inter-image homography which may be computed directly from point correspondences. Note that for a camera with fixed intrinsic parameters $\mathtt{H}_\infty$ is a conjugate of a rotation matrix and self-calibration is straightforward using Hartley's method [4].

Since $\mathtt{R}_{ij} = \mathtt{K}_j^{-1}\mathtt{H}_{ij}\mathtt{K}_i$ is a rotation matrix, it satisfies the property that $\mathtt{R}$=$\mathtt{R}^{-\top}$, leading to

$$\mathtt{K}_j{}^\top\mathtt{H}_{ij}^{-\top}\mathtt{K}_i^{-\top} = \mathtt{K}_j^{-1}\mathtt{H}_{ij}\mathtt{K}_i \quad \Rightarrow \quad \left(\mathtt{K}_j\mathtt{K}_j{}^\top\right) = \mathtt{H}_{ij}\left(\mathtt{K}_i\mathtt{K}_i{}^\top\right)\mathtt{H}_{ij}{}^\top.$$

Noting that $\mathtt{K}_i\mathtt{K}_i{}^\top$ is the dual of the image of the absolute conic (DIAC) $\boldsymbol{\omega}_i^*$ we may write

$$\boldsymbol{\omega}_j^* = \mathtt{H}_{ij}\boldsymbol{\omega}_i^*\mathtt{H}_{ij}{}^\top \tag{7}$$

which encodes the *infinite homography constraint* in the case of a stationary rotating camera with varying intrinsic parameters. This equation imposes a constraint on the transformation of the DIAC between frames and constitues the basis of our approach to self-calibration.

3.1 Self-calibration using the degenerate dual space disc quadric

In [15], Triggs introduced a clean way of expressing the self-calibration problem (in his case for a moving camera with fixed but unknown intrinsics) in terms of constraints on a quadric in $\mathcal{P}_3$ which is invariant under Euclidean transformations. In this section we show that for a rotating camera with varying intrinsics, we can derive a similar constraint, and that it is equivalent to the infinite homography constraint.

The quadric in question is the degenerate dual space disc quadric whose rim is the absolute conic in the plane at infinity. It is a projective object in 3D space which encodes metric structure and which is easier to use than the absolute conic. The representation of the quadric in a Euclidean frame is given by the rank-3 4×4 symmetric matrix:

$$\mathtt{Q}_\infty^* = \begin{bmatrix} \mathtt{I} & 0 \\ 0 & 0 \end{bmatrix} \tag{8}$$

It is easy to verify that any Euclidean transformation $\mathtt{T}$ maps $\mathtt{Q}_\infty^*$ onto itself: $\mathtt{T}\mathtt{Q}_\infty^*\mathtt{T}^\top = \mathtt{Q}_\infty^*$. The self-calibration method comprises locating the quadric in an initial projective frame and then using it to recover the projective to Euclidean transformation for the structure. $\mathtt{Q}_\infty^*$ is recovered using its *projection constraint*: $\mathtt{Q}_\infty^*$ projects onto the dual of the image of the absolute conic (DIAC)

$$\boldsymbol{\omega}_i^* = \mathtt{K}_i{\mathtt{K}_i}^\top = \mathtt{P}_i\mathtt{Q}_\infty^*{\mathtt{P}_i}^\top \tag{9}$$

independently of the projective basis chosen to express the projection matrices $\mathtt{P}_i$.

While Triggs introduced this constraint in the context of self-calibration of a moving camera with fixed intrinsics [15], Pollefeys *et al.* have recently extended the method to the case where the camera parameters may vary [10]. We now derive the *projection constraint of* $\mathtt{Q}_\infty^*$ for the case of a stationary rotating camera with varying intrinsic parameters.

Without loss of generality we may choose the first frame to be the projective basis in which the camera matrices are expressed. Therefore

$$\mathtt{P}_0 = [\mathtt{I}|0]\,, \quad \mathtt{P}_i = [\mathtt{H}_i|0]\,, \quad \mathtt{Q}_\infty^* = \begin{bmatrix} \mathtt{K}_0{\mathtt{K}_0}^\top & \mathtt{K}_0\mathbf{a} \\ \mathbf{a}^\top\mathtt{K}_0 & \mathbf{a}^\top\mathbf{a} \end{bmatrix} \tag{10}$$

where we define $\mathtt{H}_i$ to be the infinite homography between views 0 and i, and we can rewrite (9) as:

$$\boldsymbol{\omega}_i^* = \mathtt{K}_i{\mathtt{K}_i}^\top = \mathtt{P}_i \begin{bmatrix} \mathtt{K}_0{\mathtt{K}_0}^\top & \mathtt{K}_0\mathbf{a} \\ \mathbf{a}^\top\mathtt{K}_0 & \mathbf{a}^\top\mathbf{a} \end{bmatrix} {\mathtt{P}_i}^\top \tag{11}$$

where $\begin{bmatrix}\mathbf{a}^\top 1\end{bmatrix}$ is a 4-vector encoding the location of the plane at infinity Π_∞.

Combining (10) and (11) the *projection constraint* becomes

$$\boldsymbol{\omega}_i^* = \mathtt{K}_i{\mathtt{K}_i}^\top = \mathtt{H}_i\mathtt{K}_0{\mathtt{K}_0}^\top{\mathtt{H}_i}^\top = \mathtt{H}_i{\boldsymbol{\omega}_0^*}^\top{\mathtt{H}_i}^\top \tag{12}$$

Thus in the case of a rotating camera the *projection constraint of* $\mathtt{Q}_\infty^*$ reduces to the *infinite homography constraint*.

4 Self-calibration method

We shall adopt an approach similar to Pollefeys *et al.* [10] using the *infinite homography constraint* (12) to solve for the camera calibration matrices $\mathtt{K}_i$ given the set of 2D projective transformations $\mathtt{H}_i$ which relate corresponding points between the views 0 and i. A minimal parameterisation has been chosen to represent the upper triangular matrices $\mathtt{K}_i$ using the 5 intrinsic parameters of the camera: α_u, r, u_0, v_0, k.

If U is the number of unknown intrinsics in the first frame, and V is the number of intrinsics which may subsequently vary, then the total number of unknowns is $U + V(n - 1)$ where n is the number of frames. A condition for a solution is therefore

$$U + V(n-1) \leq P(n-1) \tag{13}$$

where P is the number of independent equations provided by (12) which is clearly less than or equal to 5. We therefore require $V < 5$ (i.e. strictly less than 5), meaning that not all the intrinsic parameters may be allowed to vary throughout the sequence and therefore some constraints on the parameters must be available. When this is the case, equation (12) may be solved and the calibration matrices $\mathtt{K}_i$ may be determined.

An approximate solution may be obtained using a non-linear least squares algorithm. In our implementation, a Levenberg-Marquardt algorithm was used, where the parameters to be computed were the unknown intrinsic parameters of each calibration matrix $\mathtt{K}_i$ and the cost function to be minimized was

$$\sum_{i=1}^{n} \| \mathtt{K}_i\mathtt{K}_i{}^\top - \mathtt{H}_i\mathtt{K}_0\mathtt{K}_0{}^\top\mathtt{H}_i{}^\top \|_F^2 \tag{14}$$

where $\mathtt{K}_i\mathtt{K}_i{}^\top$ and $\mathtt{H}_i\mathtt{K}_0\mathtt{K}_0{}^\top\mathtt{H}_i{}^\top$ were normalised so that their Frobenius norms were equal to one to eliminate the unknown scale factor.

The interesting property of this self-calibration method is that all the constraints available on the camera intrinsic parameters may be readily included in the model and used to constrain the minimization process since the parameterization used for the calibration matrices $\mathtt{K}_i$ explicitly uses the intrinsic parameters of the camera.

In particular, standard video cameras generally satisfy that there is no skew between the sensor array axes and that the aspect ratio is fixed. Often it may also be assumed that the principal point is located in the centre of the image and that its location does not vary significantly, so it can be assumed to be fixed.

Once the calibration matrices have been determined it is straightforward to compute the rotation matrices $\mathtt{R}_i$ which express the relative orientation of each frame with respect to the reference frame using the expression $\mathtt{R}_i = \mathtt{K}_i^{-1}\mathtt{H}_i\mathtt{K}_0$.

5 Experiments

In this section we present experimental results using our calibration method. Experiments were run both on synthetic and real sequences. The synthetic experiments proved the feasibility of the method. However, we only report the results obtained using real imagery here since ground truth data was available for comparison and they prove the self-calibration method to be very useful in practice.

5.1 Wembley sequence

Figure 1 shows a sequence of images of Wembley Stadium taken with a tripod mounted camcorder from the broadcast gantry. The camera was panned (but not tilted) and zoomed out during the sequence. The inter-image homographies have been computed using optic flow [12].

For this sequence, the aspect ratio was fixed at 1.0 and the skew fixed at 0.0, typical values for standard video equipment. The principal point was allowed to vary, since in cheaper zoom lenses the principal point tends to "barrel" as the lens is zoomed in and out, a result of minor mechanical misalignments. This phenomenon can be observed in the computed values of the principal point shown along with the computed focal lengths in figure 2. The computed rotation of the camera is plotted in figure 2. Although the true data in this case were unavailable, the computed data tally well with our expectations.

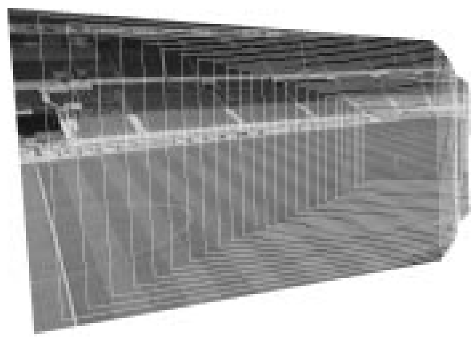

Figure 1: The mosaic constructed from a panoramic sequence of Wembley Stadium.

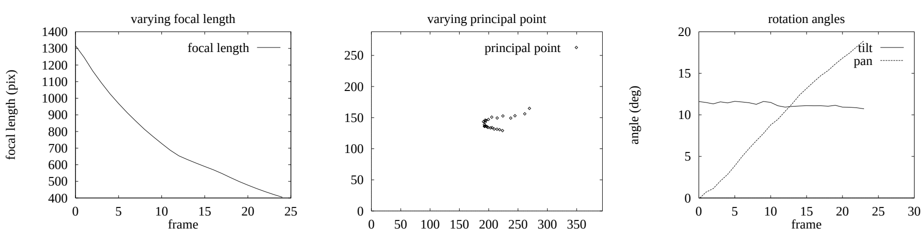

Figure 2: Values computed for the focal length (left) and principal point (centre) and pan and tilt angles (right) for the Wembley Stadium sequence.

5.2 Bookshelf sequences with ground truth data

Two image sequences were taken using a camera with a zoom lens mounted on our Yorick stereo head/eye platform [13]. In these experiments we use only two of the degrees of freedom (of the four available), using one of the two independent vergence axes to pan the camera, and the common elevation axis to tilt it. In the first sequence, the focal length of the camera remained fixed, while the pan and the tilt of the camera were varied following a circular trajectory. This experiment was carried out to assess the performance of the self-calibration method in the case of constant intrinsic parameters. In the second sequence, the focal length of the camera was set to increase linearly, using the controlled zoom lens, while the camera performed the same circular movement.

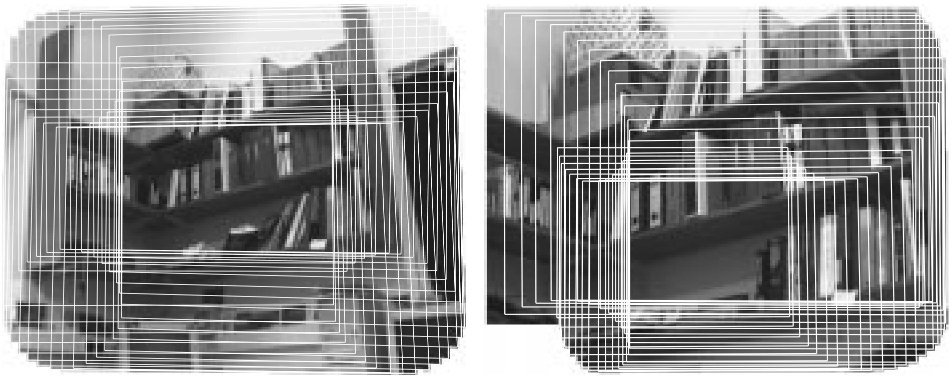

Figure 3: Mosaics constructed from the two bookshelf sequences during which the camera panned and tilted while the focal length remained fixed (left) and was varied (right).

Ground truth values for the pan and tilt angles of the camera were provided by the encoders of the head/eye platform, which are accurate to 0.01 of a degree. The servo lens provided ground truth data of the position of the zoom lens for each frame in the image sequence. The camera was then calibrated, using an accurately machined calibration grid and classical calibration algorithm, to obtain ground truth values for the internal parameters at each of the different positions of the zoom lens. Radial lens distortion was modelled using a one parameter model and the images appropriately warped to correct for this factor.

The homographies that relate corresponding points between views were computed in two stages. First, the inter-image homographies were computed from corresponding corners (detected and matched automatically). Second, the homographies were refined by minimizing the reprojection image error using a bundle-adjustment technique [1]. This second stage is usually essential in order to obtain accurate calibration results. Figure 3 shows the mosaics of both image sequences. In both sequences the aspect ratio was fixed at 1.0 and the skew at 0.0 in the self-calibration process.

Figure 4 shows the results obtained for the calibration parameters and the rotation angles, along with the ground truth data, for the constant focal length sequence. The experiment was first run assuming both the principal point and focal length to be unknown but fixed, reproducing Hartley's experiment[4], giving very accurate results. The additional experiments have (i) the focal length estimate allowed to vary but the principal point fixed, (ii) the principal point estimate allowed to vary but the focal length fixed, and (iii) both focal length and principal point estimates allowed to vary. The results match the ground truth data very accurately, except when the principal point was free to vary, in which case the focal length was underestimated.

In figure 5 we show the ground truth and computed values of the internal parameters and the rotation angles for two different experiments run on the variable focal length sequence. The principal point was assumed to be unknown but fixed in the first and allowed to vary in the second. The results obtained using the fixed principal point model match the real data very accurately. However, when the principal point was allowed to vary the focal length was overestimated by 10%, while the pan and tilt angles were still computed very accurately.

Convergence to the same values was achieved over a very wide range of starting values in both sequences.

The errors when the principal point was allowed to vary require further investigation. It seems likely that the cause is overfitting of the data, since they were acquired with a high quality zoom lens in which the true principal point varied little over the zoom range (figures 4 and 5 include the ground truth values obtained from classical calibration).

6 Discussion

We have presented a method for self-calibration of a camera which can rotate and change its intrinsics, but not translate. The basis of the method draws on ideas from several previously published self-calibration methods, in particular [4, 10], and fills one of the remaining holes in the larger self-calibration picture. We conducted experiments with real imagery and ground truth data for comparison which assess the accuracy and the stability of the self-calibration process and prove it a very useful method in practice.

Our method currently uses the inter-image homographies as input. An alternative approach we are currently investigating is more direct, namely minimising the objective $\sum_{ij} ||\mathbf{x}_i^j - \mathtt{K}_j \mathtt{R}_j \hat{\mathbf{x}}_i||$ over the calibration and rotational parameters in each frame j and over estimates of the true points on Π_∞, $\hat{\mathbf{x}}_i$ (i.e. direction vectors). An attractive feature of this method would be the possiblity of incorporating a radial correction parameter in the minimisation, $\sum_{ij} ||\mathbf{x}_i^j - g(\boldsymbol{\alpha}_j, \hat{\mathbf{x}}_i)||$, where g is a non-linear function of intrinsic and extrinsic calibration parameters $\boldsymbol{\alpha}_j$, including a term for radial distortion.

Acknowledgements

We are very grateful to our colleagues David Capel, Andrew Fitzgibbon, Torfi Thorhallsson and Andrew Zisserman. Financial support has been provided by EPSRC (ARF to IDR, and GR/L58668), EU (Marie Curie Fellowship to LA) and Research Coucil of Norway (EH).

References

[1] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proc. CVPR*, 1998.

[2] F. Du and M. Brady. Self-calibration of the intrinsic parameters of camera for active vision systems. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1993.

[3] O. D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *Intl. J. of Robotics Research*, 5(3):27–52, 1992.

[4] R. I. Hartley. Self-calibration from multiple views with a rotating camera. In *Proc. 3rd European Conf. on Computer Vision, Stockholm*, volume 1, pages 471–478, 1994.

[5] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proc. ICPR*, 1996.

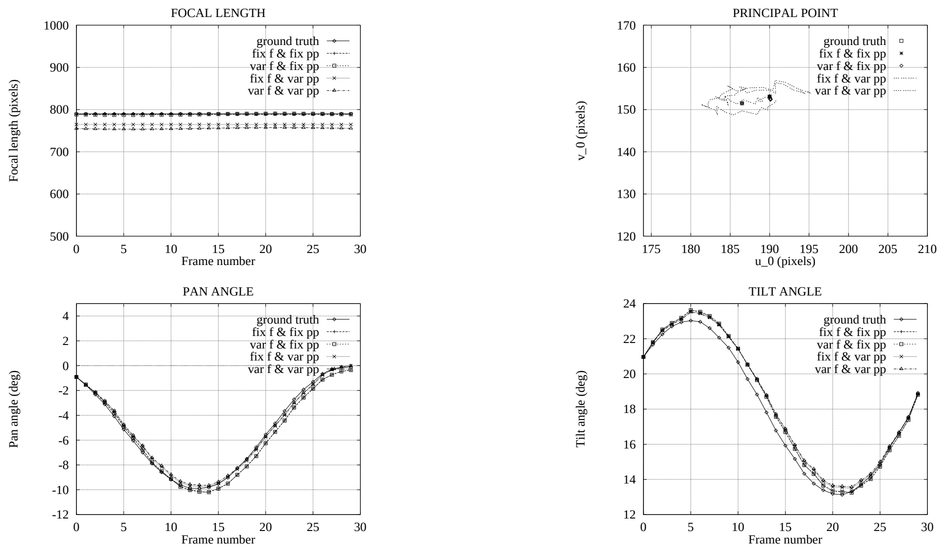

Figure 4: Ground truth and computed values for the focal length (top left) and the principal point (top right); the pan (bottom left) and the tilt (bottom right) angles of the camera for a bookshelf sequence when the focal length was constant. Note that, for clarity, the plot depicting the principal point does not show the entire image but only the central section.

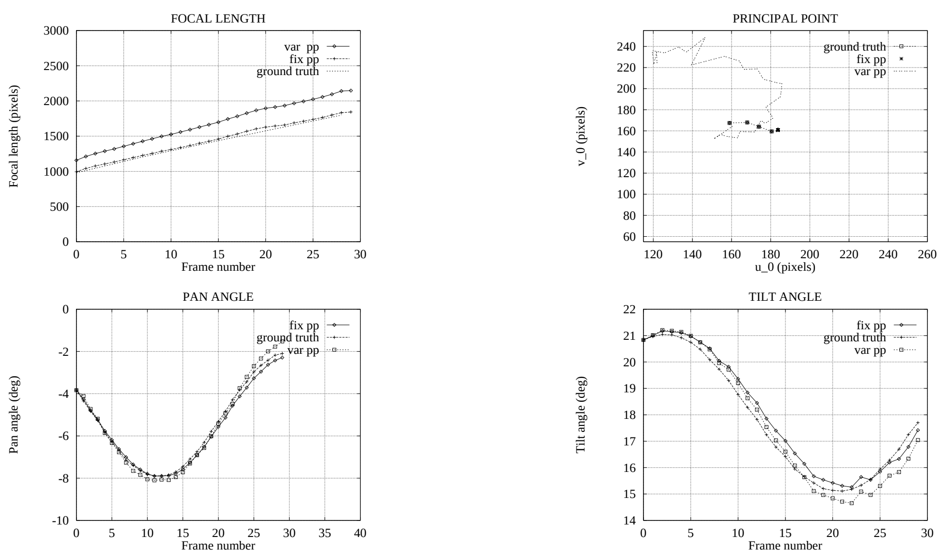

Figure 5: Ground truth and computed values for the focal length (top left) and the principal point (top right); the pan (bottom left) and the tilt (bottom right) angles of the camera during a bookshelf sequence while zooming. For clarity, only the central section of the image was depicted in the plot showing the principal point.

[6] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. CVPR*, 1997.

[7] Q.-T. Luong and T. Viéville. Canonic representations for the geometries of multiple projective views. In *Proc. 3rd European Conf. on Computer Vision, Stockholm*, pages 589–599, May 1994.

[8] S.J. Maybank and O. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[9] P.F. McLauchlan and D.W. Murray. Active camera calibration for a Head-Eye platform using the variable State-Dimension filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(1):15–22, 1996.

[10] M. Pollefeys, R. Koch, and L. Van Gool. Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. ICCV*, pages 90–96, 1998.

[11] M. Pollefeys, L. Van Gool, and A. Oosterlinck. The modulus constraint: a new constraint for self-calibration. In *Proc. ICPR*, pages 31–42, 1996.

[12] I. Reid and A. North. 3D trajectories from a single viewpoint using shadows. In *Proc. BMVC*, 1998.

[13] P. M. Sharkey, D. W. Murray, S. Vandevelde, I. D. Reid, and P. F. McLauchlan. A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4):517–535, 1993.

[14] G. Stein. Accurate internal camera calibration using rotation, with analysis of sources of error. In *Proc. 5th Int'l Conf. on Computer Vision, Boston*, pages 230–236, 1995.

[15] W. Triggs. Auto-calibration and the absolute quadric. In *Proc. CVPR*, pages 609–614, 1997.

[16] A. Zisserman, P. Beardsley, and I. Reid. Metric calibration of a stereo rig. In *IEEE Workshop on Representation of Visual Scenes, Boston*, pages 93–100, 1995.