[Journal Name], [Volumn Number], 1–23 ([Volumn Year])

Self-calibration of rotating and zooming cameras

L. DE AGAPITO*, E. HAYMAN** AND I. REID
Department of Engineering Science, Oxford University
Parks Road, Oxford, OX1 3PJ, UK

Received ??; Revised ??

Abstract. In this paper we describe the theory and practice of self-calibration of cameras which are fixed in location and may freely rotate while changing their internal parameters by zooming. The basis of our approach is to make use of the so-called *infinite homography constraint* which relates the unknown calibration matrices to the computed inter-image homographies. In order for the calibration to be possible some constraints must be placed on the internal parameters of the camera.

We present various self-calibration methods. First an iterative non-linear method is described which is very versatile in terms of the constraints that may be imposed on the camera calibration: each of the camera parameters may be assumed to be known, constant throughout the sequence but unknown, or free to vary. Secondly, we describe a fast linear method which works under the minimal assumption of zero camera skew or the more restrictive conditions of square pixels (zero skew and known aspect ratio) or known principal point. We show experimental results on both synthetic and real image sequences (where ground truth data was available) to assess the accuracy and the stability of the algorithms and to compare the result of applying different constraints on the camera parameters. We also derive an optimal Maximum Likelihood estimator for the calibration and the motion parameters. Prior knowledge about the distribution of the estimated parameters (such as the location of the principal point) may also be incorporated via Maximum a Posteriori estimation.

We then identify some near-ambiguities that arise under rotational motions showing that coupled changes of certain parameters are barely observable making them indistinguishable. Finally we study the negative effect of radial distortion in the self-calibration process and point out some possible solutions to it.

1. Introduction

We are concerned in this paper with the problem of determining the internal parameters of a camera which is free to rotate and zoom, but which remains in one location, using unstructured visual data. The internal parameters of a camera determine the mapping from image locations to rays in 3D Euclidean space, and for this reason so-called "camera calibration" has been the subject of research right from the start of the short history of machine vision.

While traditional approaches (see, e.g. [Tsai and Lenz1989]) made use of known scene structure such as accurate grids, or accurately surveyed beacons, relatively recently 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 [Faugeras et al.1992], which showed that self-calibration was theoretically and practically feasible for a camera moving through an unknown scene with constant but unknown intrinsics. Since

*Now at Queen Mary, University of London (U.K.)
**Now at the Royal Institute of Technology KTH (Sweden)

that time various methods have been developed [Hartley1994, Luong and Viéville1996, Heyden and Åström1996, Triggs1997] including some to deal with degenerate types of motion such as pure translation [Moons et al.1993] or pure rotation [Hartley1997].

More recently, the observation was made in [Heyden and Åström1997, Pollefeys et al.1998, Heyden and Åström1999, Hartley et al.1999] that self calibration is possible under much looser assumptions; some, but not all of the camera intrinsics may vary. This extension of the theory of self-calibration allows calibration to be carried out on video sequences with a zooming camera.

In this paper we explore the case of a stationary camera which may rotate and change its intrinsics, a situation 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. 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 [Hartley1997] and Pollefeys *et al.* [Pollefeys et al.1998], but differs from the former in that we consider the case of varying rather than constant intrinsics, and from the latter in that we consider pure rotations, a case not handled by that work. It can also be related to work on self-calibration from planar scenes for cameras undergoing general motion [Triggs1998, Liebowitz and Zisserman1998, Sturm and Maybank1999, Zhang1999]. However, Triggs' method applies to a camera with constant intrinsics, Liebowitz and Zisserman require some metric information about the world plane to be known and the works by Sturm and Maybank and Zhang require a plane with known structure, whereas our methods deal with the special case of the plane being located at infinity.

In this paper we elaborate on, and extend our own previous work in this area [Agapito et al.1998, Agapito et al.1999, Agapito et al.2001] in a number of ways. In addition to a detailed description of that work, we:

- consider optimal estimation of the calibration parameters;
- examine some near-ambiguities which arise in poorly conditioned sequences (small motions and large focal lengths) and do not permit the simultaneous computation of the motion parameters and the internal parameters of the camera.
- explore the issue of calibration of non-linear effects such as radial lens distortion;
- demonstrate applications of the theoretical results.

Recent work in self-calibration of rotating and zooming cameras also includes work by Seo and Hong [Seo and Hong1998, Seo and Hong1999].

The paper is organised as follows. We begin with background material, describing pinhole projection equations and discussing the specific case of zero translation (section 2.1). We then develop a constraint which is the basis for self-calibration (section 2.3). Two methods for self calibration, a non-linear and a linear one are considered (section 3) and we present experimental results for a variety of experiments with both synthetic and real data (section 4). We go on to consider optimal parameter estimation (section 5), near-ambiguities (section 6) and the effects of radial distortion (section 7). An application of the work to building spherical and cylindrical mosaics is demonstrated (section 8) and we conclude with a discussion of the work and future directions (section 9).

2. The rotating camera

2.1. Camera model

The projection of scene points onto an image by a perspective camera may be modelled by the central projection equation $\mathbf{x} = \mathtt{P}\mathbf{X}$, where $\mathbf{x} = [x\ y\ 1]^\top$ 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. Note that this equation holds only up to scale. The matrix $\mathtt{P}$ is a rank-3 matrix which may be decomposed as $\mathtt{P} = \mathtt{K}\,[\mathtt{R}|\mathbf{t}]$, where the rotation $\mathtt{R}$ and the translation $\mathbf{t}$ represent the Euclidean transformation between the camera and the world coordinate systems and $\mathtt{K}$ is an upper triangular matrix which

encodes the internal parameters of the camera in the form

$$\mathtt{K} = \begin{pmatrix} \gamma f & sf & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

Here, f is the focal length and γ is the aspect ratio The principal point is (u_0, v_0) and s is a skew parameter which is a function of the angle between the horizontal and vertical axes of the sensor array.

We choose, without loss of generality, to locate the origin of the camera's coordinate system at the optic centre. Hence the projection matrix for each view i may be written as

$$\mathtt{P}_i = \mathtt{K}_i \left[\mathtt{R}_i | 0\right] \tag{2}$$

The projection of a scene point $\mathbf{X} = [X\ Y\ Z\ 1]^\top$ onto an image point $\mathbf{x}$ may now be expressed as

$$\mathbf{x} = \mathtt{K}_i \left[\mathtt{R}_i | 0\right] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \mathtt{K}_i \mathtt{R}_i \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathtt{K}_i \mathtt{R}_i \overline{\mathbf{X}} \tag{3}$$

Since the last column of the projection matrix is zero, the depth of the world points along the ray is irrelevant and we only consider the projection of 3D rays $\overline{\mathbf{X}}$. Therefore, in the case of a rotating camera, the mapping of 3D rays to image points is encoded by the 3×3 invertible projective transformation

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

2.2. The inter-image homography

Given a 3D ray $\overline{\mathbf{X}}$, its projections onto two different images will be

$$\mathbf{x}_i = \mathtt{K}_i \mathtt{R}_i \overline{\mathbf{X}} \tag{5}$$

$$\mathbf{x}_j = \mathtt{K}_j \mathtt{R}_j \overline{\mathbf{X}} \tag{6}$$

Eliminating $\overline{\mathbf{X}}$ from these equations it is easy to see that in the case of a rotating camera there exists a global 2D projective transformation (homography) $\mathtt{H}_{ij}$ that relates corresponding points in two views:

$$\mathbf{x}_j = \mathtt{H}_{ij} \mathbf{x}_i \tag{7}$$

whose analytic expression is given by:

$$\mathtt{H}_{ij} = \mathtt{K}_j \mathtt{R}_j \mathtt{R}_i^{-1} \mathtt{K}_i^{-1} = \mathtt{K}_j \mathtt{R}_{ij} \mathtt{K}_i^{-1} \tag{8}$$

The inter-image homographies $\mathtt{H}_{ij}$ may be calculated directly from image measurements, for instance from point or line correspondences, or direct methods based on image intensity.

2.3. The infinite homography constraint

We will now derive the constraint that relates the homographies to the calibration matrices for each view. 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

$$\mathtt{K}_j{}^\top \mathtt{H}_{ij}^{-\top} \mathtt{K}_i^{-\top} = \mathtt{K}_j^{-1} \mathtt{H}_{ij} \mathtt{K}_i \tag{9}$$

and

$$\left(\mathtt{K}_j \mathtt{K}_j{}^\top\right) = \mathtt{H}_{ij} \left(\mathtt{K}_i \mathtt{K}_i{}^\top\right) \mathtt{H}_{ij}{}^\top \tag{10}$$

This same expression is also obtained by projecting a point on the plane at infinity, $\mathbf{X} = [X\ Y\ Z\ 0]^\top$, onto a camera with a non-zero fourth column in $\mathtt{P}_i$. The observed inter-image homographies $\mathtt{H}_{ij}$ are thus the homographies induced by the plane at infinity i.e. the infinite homographies $\mathtt{H}_\infty$.

Thus equation (10) is known as the *infinite homography constraint*; it relates the camera calibration matrices to the infinite homographies and constitutes the constraint we will use for self-calibration.

This constraint may also be interpreted geometrically in terms of the absolute conic and its projection on the image plane. The absolute conic Ω_∞ is a point conic in 3 space which is invariant to Euclidean transformations. It consists of points $\mathbf{X} = [X\ Y\ Z\ 0]^\top$ on the plane at infinity such that $X^2 + Y^2 + Z^2 = 0$ or alternatively $\mathbf{X}^\top \mathbf{X} = 0$. This implies that its expression in the Euclidean frame is given by the identity matrix $\mathtt{I}$.

Since points on the absolute conic are on the plane at infinity they project onto the image plane following the expression $\mathbf{x} = \mathtt{K}\mathtt{R}\overline{\mathbf{X}}$, where $\overline{\mathbf{X}} = [X\ Y\ Z]^\top$. Therefore, points on the image of the absolute conic (IAC) $\boldsymbol{\omega}$ must satisfy: $\mathbf{x}^\top (\mathtt{K}\mathtt{K}^\top)^{-1} \mathbf{x} = 0$ and the expression of the IAC is thus given by $\boldsymbol{\omega} = (\mathtt{K}\mathtt{K}^\top)^{-1}$. What is important to note here is that the IAC only depends on the cal-

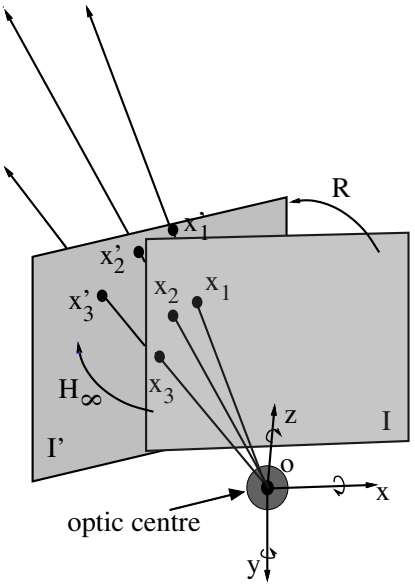


Fig. 1. Corresponding image points are related by the infinite homography $\mathtt{H}_\infty$.

ibration parameters of the camera, therefore determining the location of the IAC is equivalent to knowing the calibration of the camera.

Noting that its inverse is the dual space line conic, also called the dual image of the absolute conic (DIAC) $\boldsymbol{\omega}^* = \mathtt{K}\mathtt{K}^\top$, we may rewrite the infinite homography constraint (10) as

$$\boldsymbol{\omega}_j^* = \mathtt{H}_{ij}\boldsymbol{\omega}_i^*{\mathtt{H}_{ij}}^\top \tag{11}$$

When the internal parameters of the camera are varying, the DIAC is different for each frame and the infinite homography constraint describes its mapping between image planes [Luong and Viéville1996]. It relates the 2D projective transformations $\mathtt{H}_{ij}$ to the calibration matrices for each frame $\mathtt{K}_i$ and will constitute the basis for the self-calibration algorithms we will describe.

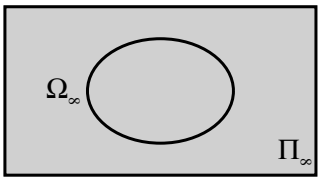
2.3.1. An alternative derivation of the infinite homography constraint A clean way of expressing the self-calibration problem was presented in [Hartley1993] and later in [Heyden and Åström1996, Triggs1997] (for the case of a moving camera with constant but unknown intrinsics) in terms of constraints on a quadric in $\mathcal{P}^3$ which is invariant under Euclidean transformations. We now 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:

$$\mathtt{Q}_\infty^* = \begin{bmatrix} \mathtt{I} & 0 \\ 0 & 0 \end{bmatrix} \tag{12}$$

It is easy to verify that any Euclidean transformation $\mathtt{T}$ maps $\mathtt{Q}_\infty^*$ to 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 image of the absolute conic (DIAC)

$$\boldsymbol{\omega}_i^* = \mathtt{K}_i{\mathtt{K}_i}^\top = \mathtt{P}_i\mathtt{Q}_\infty^*{\mathtt{P}_i}^\top \tag{13}$$



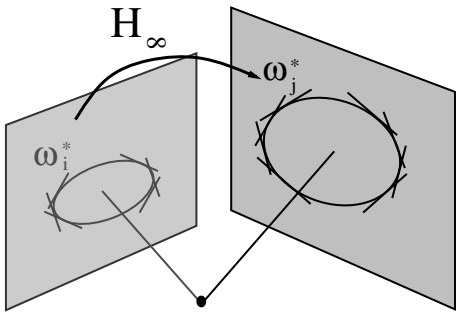


Fig. 2. The absolute conic Ω_∞ is a point conic that lies on the plane at infinity Π_∞ which is invariant to Euclidean transformations. It projects onto the *image of the absolute conic* (IAC) ω, which depends only on the internal parameters of the camera at each frame. Its dual space line conic is the *dual image of the absolute conic* (DIAC) ω^*. The infinite homography $\mathtt{H}_\infty$ maps the DIAC ω^* between views: $\boldsymbol{\omega}_j^* = \mathtt{H}_\infty\boldsymbol{\omega}_i^*{\mathtt{H}_\infty}^\top$.

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

While this constraint was originally introduced in the context of self-calibration of a moving camera with constant intrinsics [Hartley1993, Heyden and Åström1996, Triggs1997], Pollefeys *et al.* extended the method to the case where the camera parameters may vary [Pollefeys et al.1998]. 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

$$\mathtt{P}_0 = [\mathtt{I}|\mathtt{0}]\,, \quad \mathtt{P}_i = [\mathtt{H}_i|0] \tag{14}$$

where we define $\mathtt{H}_i$ to be the infinite homography between views 0 and i, and

$$\mathtt{Q}^*_\infty = \begin{bmatrix} \mathtt{K}_0\mathtt{K}_0{}^\top & -\mathtt{K}_0\mathtt{K}_0^\top\mathbf{a} \\ -\mathbf{a}^\top\mathtt{K}_0\mathtt{K}_0^\top & \mathbf{a}^\top\mathtt{K}_0\mathtt{K}_0^\top\mathbf{a} \end{bmatrix} \tag{15}$$

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

We can rewrite (13) as:

$$\boldsymbol{\omega}^*_i = \mathtt{K}_i\mathtt{K}_i{}^\top = \mathtt{P}_i \begin{bmatrix} \mathtt{K}_0\mathtt{K}_0{}^\top & -\mathtt{K}_0\mathtt{K}_0^\top\mathbf{a} \\ -\mathbf{a}^\top\mathtt{K}_0\mathtt{K}_0^\top & \mathbf{a}^\top\mathtt{K}_0\mathtt{K}_0^\top\mathbf{a} \end{bmatrix} \mathtt{P}_i{}^\top \tag{16}$$

The *projection constraint* becomes

$$\boldsymbol{\omega}^*_i = \mathtt{K}_i\mathtt{K}_i{}^\top = \mathtt{H}_i\mathtt{K}_0\mathtt{K}_0{}^\top\mathtt{H}_i{}^\top = \mathtt{H}_i\boldsymbol{\omega}^*_0\mathtt{H}_i{}^\top \tag{17}$$

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

3. Self-calibration of rotating cameras

3.1. Problem formulation

The problem of self-calibrating a rotating and zooming camera is that of determining the calibration matrices for each frame, given only correspondences between views.

The infinite homography constraint

$$\mathtt{K}_j\mathtt{K}_j{}^\top = \mathtt{H}_{ij}\mathtt{K}_i\mathtt{K}_i{}^\top\mathtt{H}_{ij}{}^\top \tag{18}$$

relates the calibration matrices $\mathtt{K}_j$ to the 2D projective transformations $\mathtt{H}_{ij}$ which may be computed directly from corresponding features between images. Given at least 4 correspondences between views the infinite homographies $\mathtt{H}_{ij}$ may be calculated and the infinite homography constraint may be used to compute the calibration matrices $\mathtt{K}_j$.

In this section we present different methods for this self-calibration.

3.2. Constant intrinsic parameters

In the case where the camera's internal parameters remain constant throughout the sequence ($\mathtt{K}_i = \mathtt{K}, i = 1...n$) the infinite homography takes the form

$$\mathtt{K}\mathtt{K}^\top = \boldsymbol{\omega}^* = \mathtt{H}_{ij}\boldsymbol{\omega}^*\mathtt{H}_{ij}{}^\top \tag{19}$$

This expression may be used to generate a set of linear equations in the entries of the *dual image of the absolute conic*, $\boldsymbol{\omega}^*$, which may be used to solve for $\boldsymbol{\omega}^*$, and subsequently for $\mathtt{K}$ by Choleski factorization. This problem was solved by Hartley in [Hartley1997].

3.3. Varying intrinsic parameters

Our main contribution has been to extend the self-calibration capability to the case where the internal parameters of the camera are not constant but change throughout the sequence. This would be useful to self-calibrate a sequence taken by a zooming camera where the focal length and perhaps other parameters are changing between views.

Here we describe two different algorithms: a non-linear algorithm which is more versatile in the nature of the constraints which can be imposed on the intrinsic parameters of the camera, and a simpler linear algorithm which involves solving a set of linear equations to find the calibration parameters but which is more restrictive in the type of constraints which can be imposed on the internal parameters of the camera.

3.3.1. Non-linear algorithm It is possible to use the *infinite homography constraint* (11) using an approach similar to Pollefeys *et al.* [Pollefeys et al.1998] to solve for the camera calibration matrices $\mathtt{K}_j$ given the set of 2D projective transformations $\mathtt{H}_{ij}$ which relate corresponding points between views. Without loss of generality we may

choose the origin of the coordinate system to be aligned with the first frame such that $\mathtt{H}_{00} = \mathtt{I}$ and $\mathtt{H}_{0j} = \mathtt{H}_j$. We may now write the *infinite homography constraint* as:

$$\mathtt{K}_j{\mathtt{K}_j}^\top = \mathtt{H}_j\mathtt{K}_0{\mathtt{K}_0}^\top{\mathtt{H}_j}^\top \tag{20}$$

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

$$U + V(n-1) \leq P(n-1) \tag{21}$$

where P is the number of independent equations provided by (20) 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 provided. When this is the case, equation (20) may be solved and the calibration matrices $\mathtt{K}_j, j = 0...n-1$ may be determined. The minimal assumption for this algorithm is that at least one parameter must be constant throughout the sequence.

A solution may be obtained using a non-linear least squares algorithm. In our implementation [Agapito et al.1998], a Levenberg-Marquardt algorithm is used, where the parameters to be computed are the unknown intrinsic parameters of each calibration matrix $\mathtt{K}_j$ and the cost function to be minimized is

$$\mathcal{D} = \sum_{j=1}^{n-1} \parallel \mathtt{K}_j{\mathtt{K}_j}^\top - \mathtt{H}_j\mathtt{K}_0{\mathtt{K}_0}^\top{\mathtt{H}_j}^\top \parallel_F^2 \tag{22}$$

where the subscript F indicates the use of the Frobenious norm, and where $\mathtt{K}_j{\mathtt{K}_j}^\top$ and $\mathtt{H}_j\mathtt{K}_0{\mathtt{K}_0}^\top{\mathtt{H}_j}^\top$ are normalised so that their Frobenius norms are equal to one to eliminate the unknown scale factor.

The calibration matrices $\mathtt{K}_j$ are then parameterized explicitly in terms of their internal parameters permitting us to impose any constraints available on the intrinsic parameters. This provides great flexibility as the camera parameters may be assumed to be known, unknown but constant or varying. In particular, one may impose standard constraints such as zero camera skew or known aspect ratio or less restrictive ones such as constant but unknown skew, aspect ratio or principal point.

Another important aspect is that we have observed that the algorithm converges to the global minumum when initialized from a wide range of starting points for the camera internal parameters. In practice we use the output from the linear algorithm described below as the starting point for the non-linear minimization, but sensible guesses of the internal parameters of the camera appear to converge to the same solution.

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

3.3.2. A linear algorithm The use of the *infinite homography constraint* (11) does not provide a simple way of imposing minimal constraints on the calibration parameters (such as zero skew or known aspect ratio) linearly. However, it was noticed in [Zisserman et al.1998] and more recently used in [Agapito et al.1999], that taking the inverse of (11) gives:

$$\boldsymbol{\omega}_j = {\mathtt{H}_{ij}}^{-\top}\boldsymbol{\omega}_i\mathtt{H}_{ij}^{-1} \tag{23}$$

where now the inverse of the infinite homography constraint on the *image of the absolute conic*, $\boldsymbol{\omega}_i = {\mathtt{K}_i}^{-\top}\mathtt{K}_i^{-1}$, leads to some linear constraints on the intrinsic parameters. One may verify that if the skew of the camera is zero ($s = 0$) the form of the *image of the absolute conic* $\boldsymbol{\omega}_j$ in each frame is:

$$\begin{aligned}\boldsymbol{\omega}_j &= \mathtt{K}^{-\top}\mathtt{K}^{-1} \\ &= \begin{bmatrix} 1/(\gamma f)^2 & 0 & -u_0/(\gamma f)^2 \\ 0 & 1/f^2 & -v_0/f^2 \\ -u_0/(\gamma f)^2 & -v_0/f^2 & 1 + u_0^2/(\gamma f)^2 + v_0^2/f^2 \end{bmatrix}\end{aligned}$$

The important thing to note here is that this assumption leads to a linear constraint in the coefficients of each $\boldsymbol{\omega}_j$: if skew is zero then the coefficient $\omega_{12} = 0$. There are additional constraints on the camera intrinsics that also lead to linear equations. We summarize these constraints here:

1. If skew is zero then $\omega_{12} = 0$.
2. If skew is zero and aspect ratio is 1 (square pixels constraint) then $\omega_{11} = \omega_{22}$.
3. If skew is zero and $u_0 = 0$, then $\omega_{13} = 0$. Similarly if $v_0 = 0$ then $\omega_{23} = 0$.

Each of these constraints will give one linear equation per frame in the entries of each $\boldsymbol{\omega}_j$. Using the infinite homography constraint (23) one may transfer the constraints to constraints on the image of the absolute conic only in the reference frame $\boldsymbol{\omega}_i$. Now each constraint will provide one linear equation in the 6 independent entries of $\boldsymbol{\omega}_i$. If more than 5 equations are available the problem may be solved via least-squares.

The obvious advantage of the linear algorithm is that it is a simple method that does not require an initial estimate, hence it will not have convergence problems. It is also a very fast method making it suitable in real time applications. However, it has the disadvantage of not incorporating some useful constraints on the calibration parameters such as unknown but constant parameters (skew, aspect ratio, principal point). An alternative to solve for such parameters would be to use the residual from the linear algorithm as the cost function in a search over the parameter space to determine the value that best fits the data.

The self-calibration algorithms we have presented in this section are also applicable in the case of cameras undergoing general motion to upgrade the calibration from affine to Euclidean space. Obtaining affine structure is equivalent to knowing the location of the plane at infinity. This in turn is equivalent to determining the infinite homographies. Therefore, once these are known the rest of the self-calibration problem is reduced to that of self-calibrating a non-translating camera and the *infinite homography constraint* may be used to determine the calibration matrix for each frame.

Indeed these self-calibration methods may also be used to guide the search for the plane at infinity. Given a choice for the location of the plane at infinity, the residual given by the self-calibration algorithm may guide the search over a range of feasible values for the plane at infinity that gives the best calibration. The result is a stratified algorithm for self calibration, applicable to cameras undergoing general motion with changing internal parameters, in which one proceeds from projective to affine to Euclidean reconstruction. This algorithm is described in [Hartley et al.1999].

4. Experimental results

4.1. Experiments with synthetic data

Experiments were first carried out on synthetic data to evaluate the performance of the linear and non-linear self-calibration algorithms using different constraints on the internal parameters. The data were created to simulate a camera with a zoom lens providing a total focal length range of 12.5 mm to 35 mm and the CCD size was 7×5 mm. A cloud of 1000 points was randomly generated within a confined cubic space of 3 m side lying in front of the rotating camera at a distance of 5 m. The points were projected onto each of the image planes arising from the different orientations of the camera and the location of each image point was then perturbed in a random direction by a distance governed by a Gaussian distribution with zero mean and standard deviation σ measured in pixels. The skew of the image axes was taken to be zero, the aspect ratio of the image pixels equal to one, and the principal point was located at the centre of the image. The camera motion was such that the principal ray described a circular trajectory, simulating the motion of a pan-tilt unit. The focal length of the camera increased linearly throughout the 30 frame sequence from a value of 800 pixels to 1400 pixels.

In Figure 6 we show the results of one run of the self-calibration algorithms on two different sequences for a typical noise level of $\sigma = 0.5$ pixel, comparing the performance of the different algorithms using different constraints on the intrinsic parameters.

The graphs compare the results of the computation of the focal length, the principal point and the motion of the camera with ground truth data. The algorithms were run on two different sequences, one where the radius of motion of the camera was larger ($\theta = 5^\circ$) and the focal length of the camera was shorter and another one with a smaller motion ($\theta = 3.5^\circ$) and larger focal length, a less well conditioned configuration.

More accurate results were obtained in both sequences when the principal point was assumed to be constant but unknown. When the location of the principal point was allowed to vary in the minimization, the effect was that, although the rate of change in focal length was correctly estimated, its

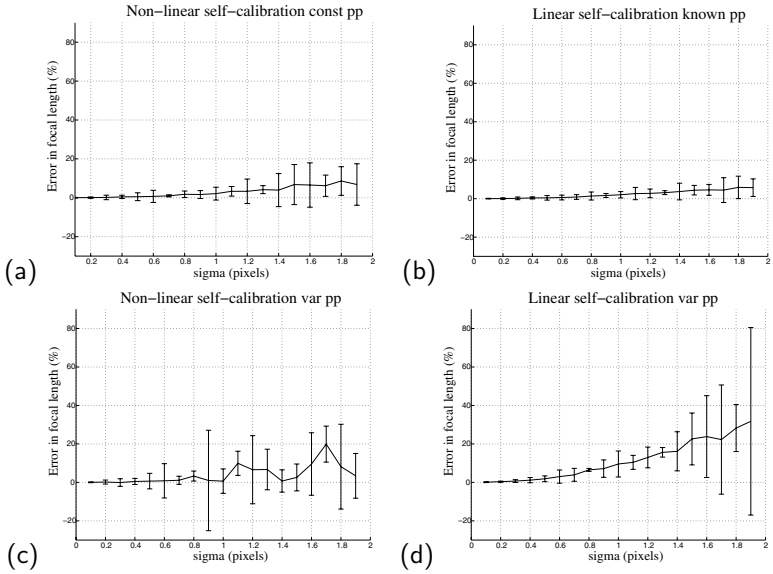


Fig. 3. This figure shows the mean and the rms error in the computation of the focal length for different noise levels. Results are represented as error bars showing the mean and lower and upper ranges of the rms relative error. The experiment was run 100 times and errors were computed over the 30 frames of the sequence. Results are shown for the *non-linear* algorithm assuming (a) constant but unknown and (b) varying principal point; and for the *linear* algorithm assuming (c) known and (d) varying principal point

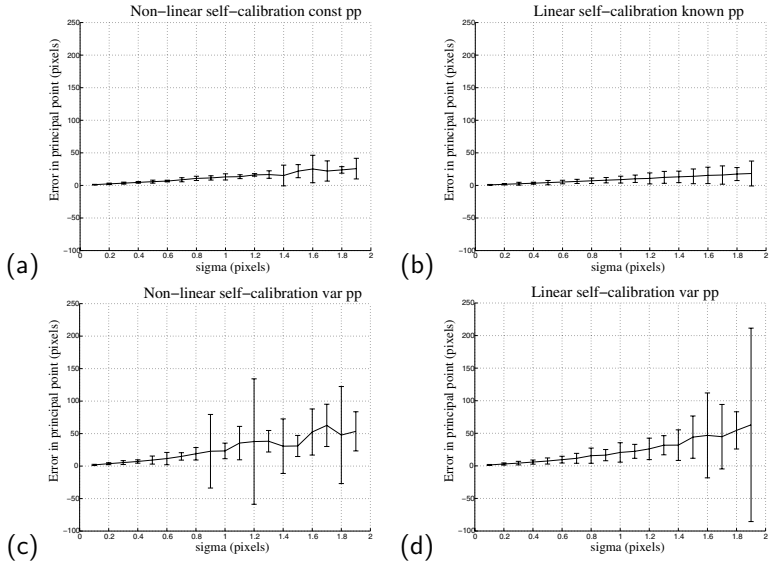


Fig. 4. This figure shows the mean and the rms error in the computation of the principal point for different noise levels. We show results from (a) the non-linear algorithm assuming constant but unknown principal point and (b) varying principal point; and from the (c) linear algorithm assuming known principal point and (d) varying principal point.

overall scale was not accurately computed. This behaviour is due to a near-ambiguity that arises in the simultaneous computation of the motion and focal length parameters and which will be described in depth in Section 6. The effect of this ambiguity is more severe in poorly conditioned sequences (small motions and large focal lengths) and when the principal point is allowed to vary.

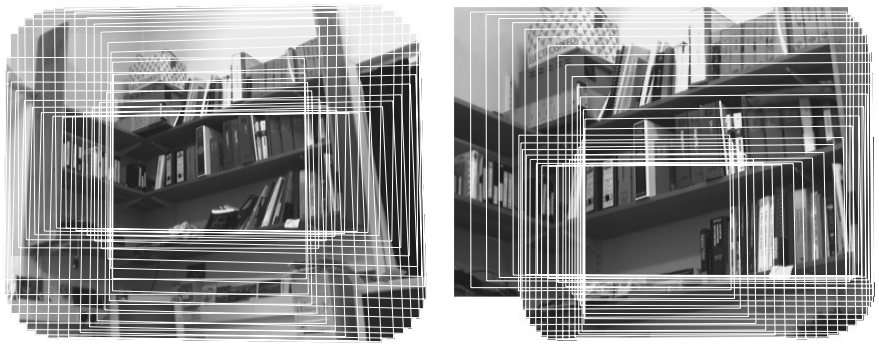


Fig. 5. Mosaics constructed from the two bookshelf sequences during which the camera panned and tilted while the focal length remained fixed (left) and was varied (right).

Figures 3 and 4 show the performance of the linear and non-linear self-calibration algorithms with respect to increasing levels of noise. The mean and relative rms error were obtained after running the experiment 100 times using different seeds to generate the noise.

The errors in computing the focal length are depicted in Figure 3 as error bars representing the mean and the upper and lower ranges of the relative rms error expressed as a percentage. These errors appear to be larger when the principal point was allowed to vary both with the non-linear and the linear algorithms. There is a noticeable bias in the estimation of the focal length when the linear method was used allowing the principal point to vary.

Figure 4 shows the mean and rms error in the computation of the location of the principal point. The error was computed as the distance of the estimated location of the principal point from its true location. Once more the error bars represent the mean and the upper and lower values of the rms error. Both the linear and non-linear methods show similar responses to noise, and errors are larger when the principal point was allowed to vary in both cases.

4.2. Experiments with real data

The image sequences used in our experiments were taken using a camera with a zoom lens mounted on our Yorick stereo head/eye platform [Sharkey et al.1993]. The camera was rotated using one of the two independent vergence axes to pan the camera, and the common elevation axis to tilt it. The mechanics of our head do not permit rotations about the Z axis. Interestingly, this situation arises very often when using stationary cameras, since they tend to be mounted on tripods with 2 degrees of freedom for rotation. As described in [Agapito et al.1999, Hayman2000] this type of motion is degenerate: imposing only the skew zero constraint fails to solve the self-calibration problem and additional constraints must be used.

Two image sequences were taken. In the first sequence, the focal length of the camera was constant, while the pan and the tilt of the camera were varied to perform a circular trajectory. In the second sequence, the focal length of the camera was set to increase linearly, using the controlled zoom lens, while the camera performed a similar circular motion. The encoders of the head/eye platform provided ground truth values for the pan and tilt angles of the camera and are accurate to 0.01 of a degree. The servo control of the zoom lens provided ground truth values of the position of the zoom lens for each frame in the image sequence. The camera was then calibrated, using an accurately manufactured calibration grid and a 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 were 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 global

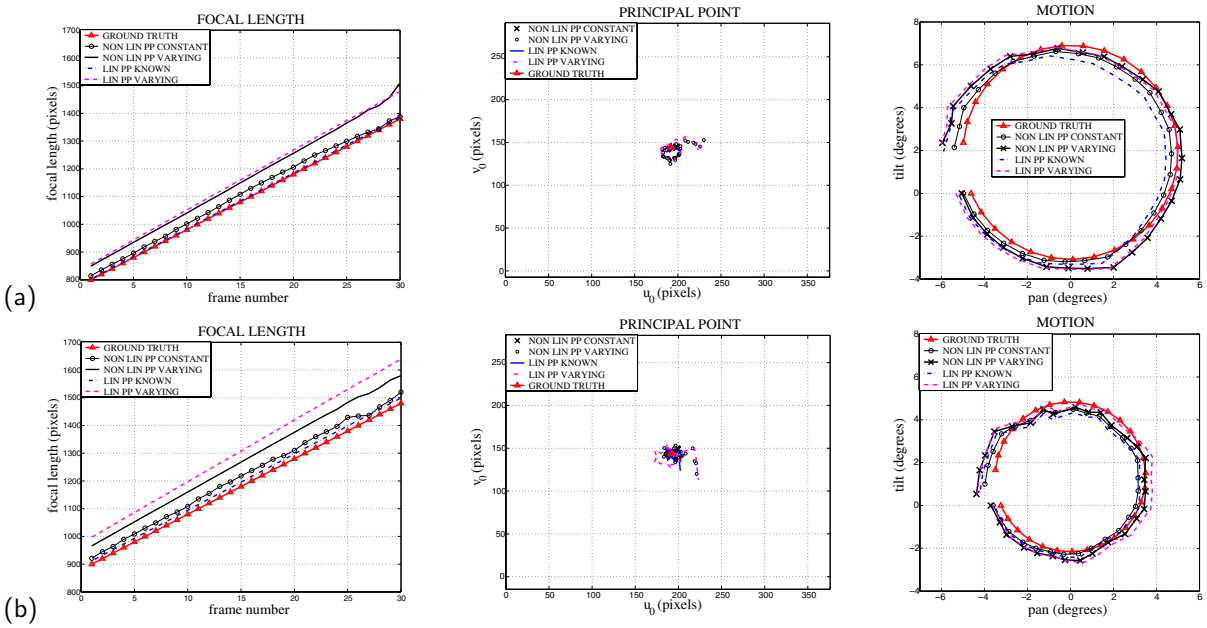


Fig. 6. Calibration results with synthetic data in the presence of image noise of $\sigma = 0.5$ pixel, showing only one run of the non-linear and linear algorithms on 2 sequences with different motions: (a) smaller focal length and motion of 5° radius and (b) larger focal length and smaller motion of 3.5° radius. We show computed values for the focal length (left), the location of the principal point (middle) and the motion (right). Results are shown for (i) the iterative Levenberg-Marquardt algorithm imposing only the square-pixels constraint (LM VARYING PP), the (ii) iterative Levenberg-Marquardt algorithm imposing both the square-pixels and the constant principal point constraints (LM CONSTANT PP), (iii) the linear algorithm imposing the square-pixels constraint (LIN VARYING PP) and (iv) the linear algorithm imposing the square-pixels and known principal point constraint (LIN KNOWN PP). For visualization purposes, the motion was represented by plotting pan versus elevation angles.

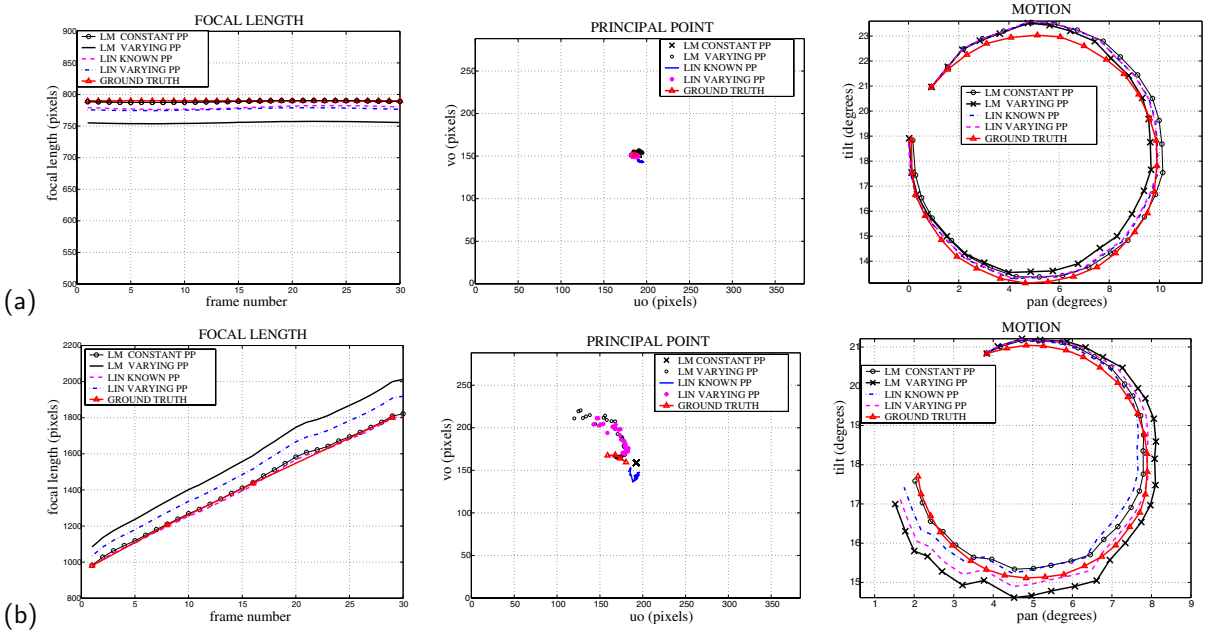


Fig. 7. Ground truth and computed values for the focal length (left), the location of the principal point (middle) and the motion of the camera (right) for the fixed focal length (top) and the variable focal length (bottom) bookshelf sequences.

image reprojection error using a bundle-adjustment technique [Capel and Zisserman1998]. This second stage is usually essential in order to obtain accurate calibration results. Figure 5 shows the mosaics of both image sequences constructed using the homographies to register the images.

Figure 7 shows results of the self-calibration algorithms applied to the constant and varying focal length sequences applying different constraints on the intrinsic parameters. The values of the aspect ratio and the skew were assumed to be known and equal to 1 and 0 respectively. Once more, the estimates of focal length and motion are greatly improved when the non-linear method is run assuming a constant but unknown principal point.

5. Optimal estimation

The previous self-calibration algorithms are not optimal in a statistical sense, since both the non-linear and linear algorithms minimize algebraic, rather than geometric errors. In this section we derive an optimal estimator for the calibration and the motion parameters for the case when point correspondences are used as input data. Similar derivations may be obtained for the case where line correspondences are used or for direct approaches.

5.1. Maximum Likelihood Estimation (MLE)

Let us consider that the noise $\mathbf{w}$ on the measured image feature positions $\hat{\mathbf{x}}$ is additive and described by a Gaussian distribution with mean zero and standard deviation σ. The measured location $\hat{\mathbf{x}}$ is thus related to the true location by:

$$\hat{\mathbf{x}} = \mathbf{x} + \mathbf{w} = \mathtt{H}(\boldsymbol{\theta}) + \mathbf{w} \tag{24}$$

where $\mathbf{x} = \mathtt{H}(\boldsymbol{\theta})$ describes the model we have for the true values of the image points given an estimate of the model parameters $\boldsymbol{\theta}$. In our case, of course, this is the projection equation:

$$\hat{\mathbf{x}} = \mathbf{x} + \mathbf{w} = \mathtt{KR}\overline{\mathbf{X}} + \mathbf{w} \tag{25}$$

It is then straightforward to prove that a Maximum Likelihood Estimate is given by:

$$\mathtt{MLE} = \arg\min_{\mathtt{K}_i\mathtt{R}_i\overline{\mathbf{X}}_j} \sum_{i=1}^{n}\sum_{j=1}^{m} \parallel \hat{\mathbf{x}}_{ij} - \mathtt{K}_i\mathtt{R}_i\overline{\mathbf{X}}_j \parallel^2 \tag{26}$$

That is, the $\mathtt{K}_i$, $\mathtt{R}_i$ and $\overline{\mathbf{X}}_j$ which minimise the sum of the squared distances of measured feature locations to the true image points for all points across all views.

The minimum of this non-linear cost function is sought using a Levenberg-Marquardt algorithm modified to take advantage of the sparse block structure of the matrices involved in the process. This method is generically termed bundle-adjustment in the computer vision and photogrammetry communities.

Given the large number of model parameters it is not surprising that the objective is highly non-convex and so it is crucial to provide an initial estimate for the iteration close to the global minimum for the algorithm to converge to it. Our overall algorithm is as follows. First compute the inter-image homographies from corresponding points between views. Then obtain an initial estimate for matrices $\mathtt{K}_i$ and $\mathtt{R}_i$ using the linear method described in section 3.3.2. Use this as the starting point for the non-linear method described in section 3.3.1. Finally, refine the estimates of the camera matrices $\mathtt{K}_i$, the rotation matrices $\mathtt{R}_i$, and the 3D rays $\overline{\mathbf{X}}_j$ using the MLE minimization.

5.2. Using priors on the estimated parameters: Maximum a Posteriori Estimation (MAP)

The principal point is known to be a poorly constrained parameter which tends to fit to noise. Thus, as we observed in our experiments, permitting it to vary indiscriminately can have very deleterious effects in the computation of the other parameters. However, the use of prior knowledge about the distribution of this parameter can be used to reduce these effects.

If we model our prior expectation that the principal point probably lies close to the centre of the image as a Gaussian distribution whose mean is the image centre, it is then easy to prove that a Maximum a Posteriori estimate of the calibration

parameters is simply:

$$\mathtt{MAP} = \arg\min_{\mathtt{K}_i\mathtt{R}_i\overline{\mathbf{X}}_j} \sum_{i=1}^{n}\sum_{j=1}^{m} \| \hat{\mathbf{x}}_{ij} - \mathtt{K}_i\mathtt{R}_i\overline{\mathbf{X}}_j\|^2 +$$

$$\sum_{i=1}^{n}(\mathbf{u}_0^i - \overline{\mathbf{u}}_0)^\top \begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_y^2\end{bmatrix}^{-1}(\mathbf{u}_0^i - \overline{\mathbf{u}}_0) \qquad (27)$$

where $\mathbf{u}_0^i = ({u_0}^i, {v_0}^i)$ is the estimate of the principal point location for each frame, σ_x and σ_y are the uncertainties in the x and y direction of the principal point and $\overline{\mathbf{u}}_0$ is its expected prior value, taken to be at the centre of the image.

That is, the new cost function is simply the cost in equation (26) with the addition of a term penalizing estimates of the principal point which are far from the image centre.

5.3. Experimental results

5.3.1. Synthetic data Experiments were run first on synthetic data. The motion of the camera and the variation of focal length was equivalent to that used in previous synthetic experiments in Section 4.1. However in this experiment the true location of the principal point varied throughout the sequence. Ground truth values and the results for the computation of the focal length, principal point and motion of the camera are shown in Figure 8. The MLE and MAP refinement algorithms were run using two different initial estimates as starting points for the minimization.

Figure 8 (a) shows the results obtained when the output of the non-linear LM algorithm using the square pixels and constant principal point constraints was used as the initial estimate for the MLE and MAP. Both MLE and MAP provided improved estimates of the calibration and motion parameters.

However, when MLE and MAP were initialized at the output of the non-linear LM method with varying principal point (see Figure 8 (b)), the estimates of the focal length and the motion improved, but were still far from the true solution. The ambiguity between focal length and rotation which will be described in Section 6 is still present and MLE and MAP do not appear to resolve it.

5.3.2. Real images Here we present the results of the MLE and MAP refinement stage on the bookshelf sequence shown in section 4.2. Once more we show results choosing two different initial estimates for the minimization. Results are shown in Figure 9 and prove the significance of choosing a good starting point for the minimization.

When the starting point was set at the output given by the non-linear algorithm imposing the square pixels and constant principal point constraints (see Figure 9 (a)), the minimization started off close to the true solution and the final estimates given by MLE and MAP were very close to the global minimum. However, when the initial estimate was chosen to be the output from the non-linear algorithm allowing the principal point to vary (see Figure 9 (b)) the minimization started far from the true solution and MLE failed to obtain estimates close to the ground truth. These results show that optimal methods can fail to resolve the near-ambiguities which will be described in Section 6.

Best results, particularly for the motion parameters, were consistently given by the MAP estimate imposing a prior on the location of the principal point.

6. Degeneracies and near-ambiguities in self-calibration

During practical use of the self-calibration techniques discussed so far in this paper it is important to be aware of situations in which the recovery of camera and motion parameters yield insufficiently accurate or even completely meaningless results. Potential sources of such problems are i) radial distortion, ii) degenerate configurations, and iii) the information for self-calibration dropping below noise levels.

While radial distortion is the topic of section 7, this section briefly reviews literature on critical configurations before identifying how the estimates of parameters deteriorate in situations which are not degenerate, but which poorly constrain the solution because perspective effects in the cameras are less prominent.

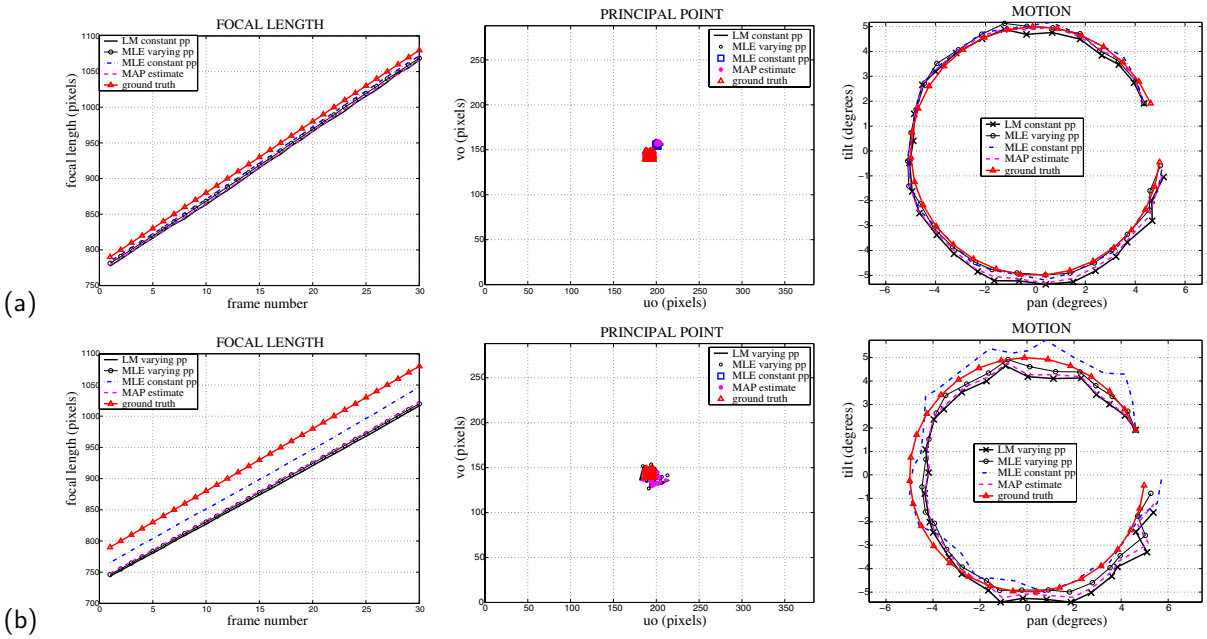


Fig. 8. Results of MLE and MAP refinement on synthetic data, where the noise level was $\sigma = 0.6$ pixels. Results are shown for one run of the self-calibration algorithm using two different starting points for the minimizations: (a) the output of the Levenberg-Marquardt (LM) minimization assuming unknown but constant principal point and (b) the output of the LM minimization allowing the principal point to vary. Graphs show (i) ground truth, (ii) starting points (LM constant/varying) (iii) MLE assuming constant principal point (MLE constant pp) and (iv) MLE with varying (MLE varying pp) principal point and (v) MAP estimation (MAP estimate).

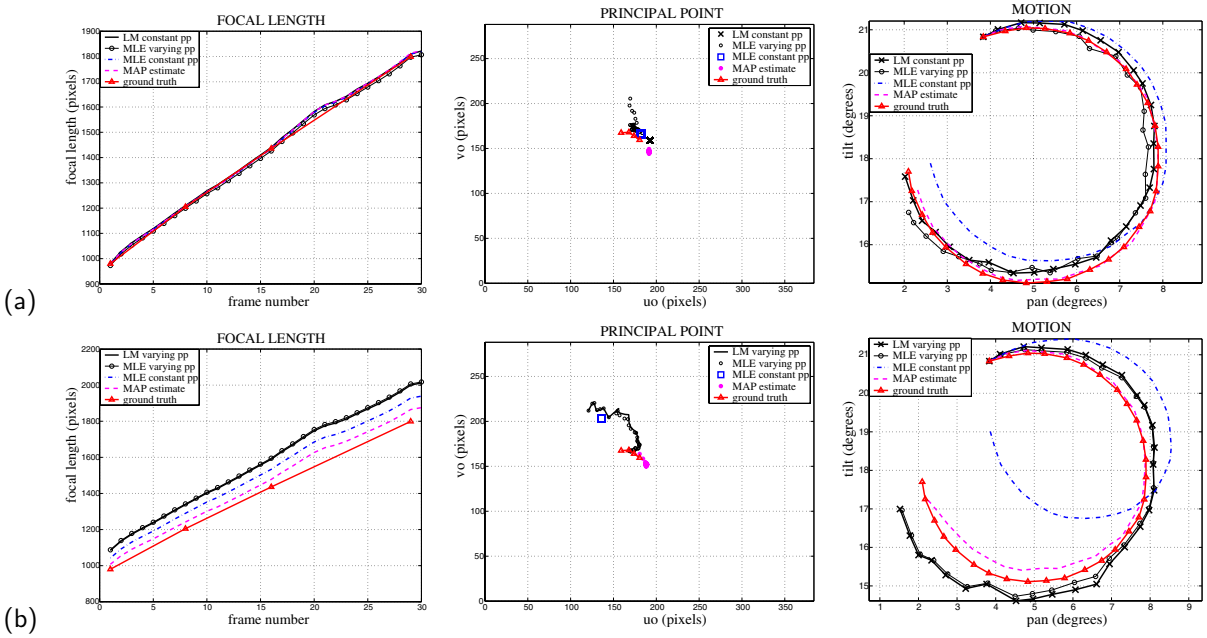


Fig. 9. Ground truth and computed values of the focal length, the location of the principal point and the motion of the camera for the variable focal length bookshelf sequence using the MLE and MAP refinement. Results are shown using different starting points. In (a) we show results using the iterative Levenberg-Marquardt algorithm imposing both the square-pixels and the constant principal point constraints (LM constant pp) as starting point, whereas in (b) the starting point was provided by the iterative Levenberg-Marquardt algorithm imposing only the square-pixels constraint, allowing the principal point to vary (LM varying pp). Results are given for (i) MLE allowing the principal point to vary (MLE varying pp), (ii) MLE with constant principal point (MLE constant pp) and (iii) maximum a posteriori estimation (MAP estimate).

6.1. Degenerate configurations: review

By a degenerate configuration, we mean that a particular motion may result on no constraints whatsoever on a given intrinsic parameter (or combination of parameters). For example, it is easy to see that a pure rotation about x- or y-axis only can never reveal anything about the camera's aspect ratio. Likewise, in Agapito *et al.* [Agapito et al.1999], the linear self calibration algorithm was used to show that the zero-skew constraint is insufficient to self-calibrate pan-tilt platforms.

Kahl, Triggs and Åström [Kahl1999, Kahl et al.2000] considered this problem, ie. of identifying multiple conics on the plane at infinity which satisfy the calibration constraints, in the context of cameras which may also translate. A valuable contribution of their work was to determine the degeneracies for single-axis rotations under various constraints on known intrinsic parameters. The case where only the focal length is unknown was also considered by Sturm [Sturm1999] and Pollefeys [Pollefeys1999]. Literature concerning the case of all intrinsic parameters being constant but unknown [Sturm1997, Zisserman et al.1998] is also relevant since if an ambiguity exists in configurations with constant intrinsics, it will certainly persist when the intrinsic parameters may change.

We do not consider degenerate configurations further in this paper, however we note that Hayman [Hayman2000] provides a full categorization of degeneracies for pan-tilt, pure pan and pure tilt platforms, building on the existence proofs of Heyden and Åström [Heyden and Åström1998, Heyden and Åström1999]. Here, the dimensionality of the family of solutions is deduced and also which parameters are and which are not recovered correctly. This work deals also with cases where one or more parameters are constant but unknown. For pure pan or tilt it demonstrates that it is insufficient to impose that the aspect ratio is constant, its value must be known, and constraints must in addition be applied to either the skew or principal point.

6.2. Near-ambiguities

We now consider two near-ambiguities present in the case of rotating cameras. Note that as opposed to the true ambiguities discussed in section 6.1, these near-ambiguities consist of coupled changes in the parameters being barely observable in the image motion and therefore often falling below noise levels.

6.2.1. Angle of rotation/focal length near-ambiguity

For small rotations there is a near-ambiguity between the rotation and the focal length, and it is difficult to distinguish between small rotations with a large focal length and larger rotations with a small focal length.

Although one may develop an argument based on small rotation approximations of homographies, the ambiguity is most easily seen by differentiating the non-homogeneous projection equation $\mathbf{x} = (f/Z)\mathbf{X}$,

$$\dot{\mathbf{x}} \;=\; \frac{f}{Z}\dot{\mathbf{X}} + \frac{\dot{f}}{Z}\mathbf{X} - \frac{f\dot{Z}}{Z^2}\mathbf{X} \qquad (28)$$

where the focal length f is a function of time and $\mathbf{x} = (x\ y\ f)$ and $\mathbf{X} = (X\ Y\ Z)$ are expressed in camera centred frames. For simplicity skew, aspect ratio and principal point are assumed known. $\dot{\mathbf{X}} = \boldsymbol{\Omega} \times \mathbf{X} + \mathbf{V}$ where $\boldsymbol{\Omega}$ is the camera's angular velocity and $\mathbf{V}$ is the velocity of the camera centre which for a purely rotating camera is zero. Substituting $\mathbf{X} = (Z/f)\mathbf{x}$ back into equation (28) then yields

$$\dot{\mathbf{x}} \;=\; \boldsymbol{\Omega} \times \mathbf{x} + \frac{\dot{f}}{f}\mathbf{x} - \frac{\hat{\mathbf{k}} \cdot \dot{\mathbf{X}}}{Z}\mathbf{x}$$

The image motion in the x-direction is thus

$$\dot{x} = f\Omega_Y - y\Omega_Z + \frac{x}{f}\left(x\Omega_Y - y\Omega_X\right) + \frac{\dot{f}}{f}x \qquad (29)$$

A similar expression holds for $\dot{y}$ of course.

Rotations about the optic axis and the *relative* change in focal length can be recovered from the terms $-y\Omega_Z$ and $\dot{f}x/f$ respectively (the latter is zoom-induced looming motion). However, the first term $f\Omega_Y$, a uniform motion of the image due to the component of rotation perpendicular to the optic axis, contains an ambiguity between focal length and rotation. The third term, $(x/f)(x\Omega_Y - y\Omega_X)$, which also arises from the component of rotation perpendicular to the optic axis, provides some disambiguating information since f and Ω_Y no longer appear as a product. Unfortunately this term is likely to disappear below noise levels for

large focal lengths (equivalently a small field of view) and also for small rotations since the term is proportional to the motion perpendicular to the optic axis. Notice too that the disambiguating information is small except at the edges of the image. Unfortunately this is also where the optical properties of the lens are poorest. Compounding these difficulties is that in practical applications, sequences taken at large f are less likely to contain significant rotation.

Consider now a sequence of several images with a small total camera rotation. Equation (29) integrates to give image motion

$$\Delta x \approx \dot{x}\Delta t = f\theta_Y - y\theta_Z + \frac{x}{f}\left(x\theta_Y - y\theta_X\right) + \frac{\Delta f}{f}x \tag{30}$$

f is then the focal length in some reference frame whereas rotation angles θ and the change in focal length Δf vary between frames. If the disambiguating term is vanishingly small, replacing f by kf, Δf by $k\Delta f$, and θ_Y by θ_Y/k in *all* frames would not change the image motion Δx. Therefore this near-ambiguity persists over a whole sequence of images. Note that k here is a constant factor over the entire sequence since f in equation (30) refers to the focal length in the reference frame.

In experiments we find that the near-ambiguity is much more pronounced when the principal point is allowed to vary in the self-calibration algorithm: with more parameters, the model is more likely to fit to the noise rather than the underlying true solution.

6.2.2. Angle of rotation/principal point near-ambiguity A similar analysis (again using the x-dimension of the image motion) shows that it is difficult to distinguish between a shift in the principal point along x and a rotation of the camera about y. Relaxing the constraint of the principal point being at (0 0), the non-homogeneous projection equation becomes $\mathbf{x} = (f/Z)\mathbf{X} + (u_0\ v_0\ 0)^\top$, and differentiation with respect to time yields

$$\dot{x} = \dot{u_0} + f\Omega_Y - y\Omega_Z + \frac{x}{f}\left(x\Omega_Y - y\Omega_X\right) + \frac{\dot{f}x}{f} \tag{31}$$

In addition to the previously mentioned near-ambiguity between focal length and motion, it is now also difficult to distinguish between motions of $f\Omega_Y$ and $\dot{u_0}$. Thus for each *single* image, if δu is the error in the estimation of the principal point in that frame, and f is the focal length, by introducing a compensatory, but erroneous rotation of $\delta u/f$ about y, the observed images are near-identical. This is illustrated in Figure 10. Again, the disambiguating information is contained in the term $(x/f)\left(x\Omega_Y - y\Omega_X\right)$, though now this has to resolve not one but two ambiguities. This explains why the ambiguity between focal length and motion is harder to resolve when the principal point is assumed to be varying between each frame in the self-calibration algorithm.

Another way of describing this near-ambiguity is that a large focal length perspective projection is hard to distinguish from a spherical projection where the principal point is meaningless. Referring to equation (31), apart from $\dot{u_0}$ the only term which describes a planar rather than spherical projection is the disambiguating term $\frac{x}{f}\left(x\Omega_Y - y\Omega_X\right)$.

6.2.3. Examples Figure 11 illustrates these near-ambiguities using both simulated and real image data. The experiments on real data use the *bookshelf* sequence, and the simulated data was synthesized similarly. Levenberg-Marquardt was used to minimize $\mathcal{D}$ in equation (22), and the principal point and focal length allowed to vary over the sequence during the minimization.

In each set of results the first two plots show the recovered and veridical focal length and principal

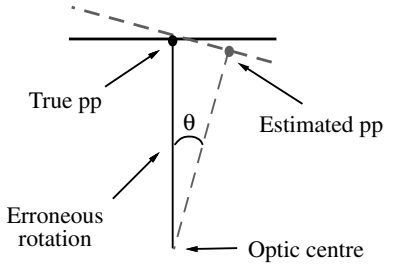


Fig. 10. The near-ambiguity between motion of the principal point and rotations. The two configurations of 1D cameras represent i) the true configuration (solid lines) and ii) a configuration where there is an error in the principal point and orientation of the camera (dashed lines). Since the two camera planes are nearly coincident, the images obtained from the two configurations are nearly identical. The problem is more pronounced the smaller the field of view of the camera.

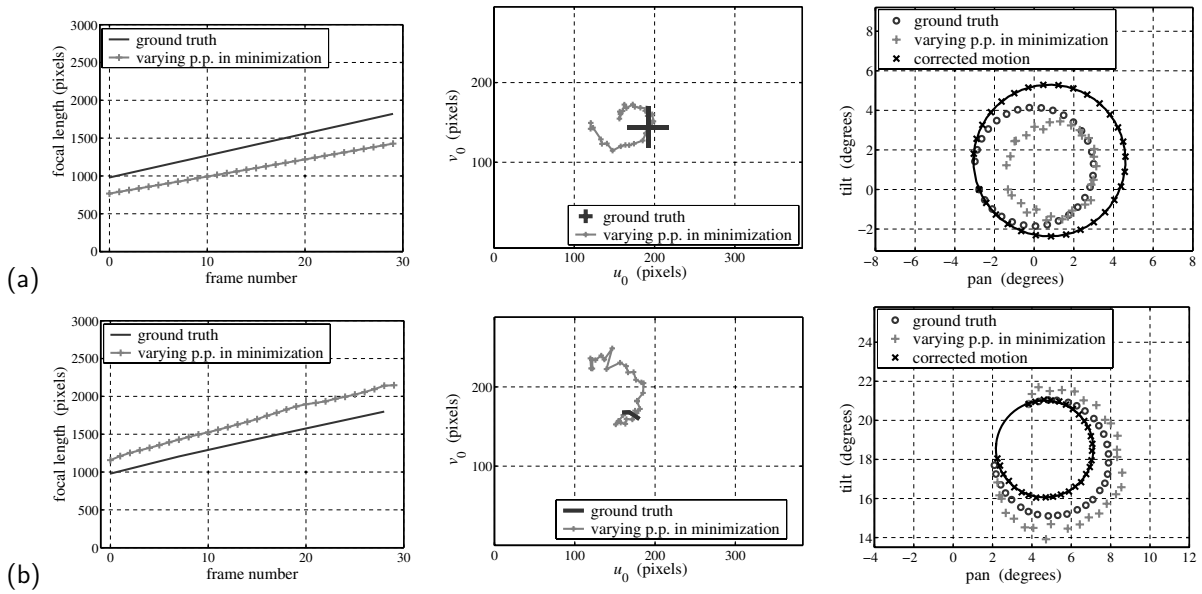


Fig. 11. Correcting the pan and tilt angles by accepting the principal-point/rotation near-ambiguity. Part (a) uses simulated data and part (b) real data (the bookshelf sequence), both use a linearly increasing focal length and motion with cone half angle 3°.

Table 1. Verification of the near-ambiguity between focal length and motion exhibited in Figure 11.

	Recovered f / true f	Radius of recovered motion / radius of true motion	Product of the two
Synthetic data, varying f	0.7819	1.2789	1.000
Real images, varying f	1.198	0.826	0.989

point. The + symbols in the third plots show the recovered camera motion in terms of pan and tilt angles. These are roughly circular, but there is a good deal of scatter about the best-fit circle.

This scatter turns out to be almost entirely due to the principal-point/rotation near-ambiguity. Using the ground truth value for the position of the principal point, the pan and tilt angles are corrected in each frame and re-plotted as × symbols. These form near perfect circles.

However, the scale of the motion is still incorrect. This is due to the near-ambiguity between focal length and rotation. Table 1 illustrates this point with the recovered scale of focal length and motion compared to the ground truth: multiplied together they give a number very close to unity.

The effect of these near-ambiguities in recovering a Euclidean reconstruction from multiple rotating cameras is described in [Hayman et al.2000] and [Hayman2000].

7. Effects of radial distortion

Radial distortion effects have frequently been neglected in computer vision, in particular in self-calibration applications. In this section we will prove that they can have a very severe effect on the self-calibration process. Thus, a proper estimation of the radial distortion parameters is necessary to achieve good self-calibration results. The radial distortion that occurs in a lens may be modelled using the second order radial transformation

$$\mathbf{r}_d = \frac{\mathbf{r}_u}{1 - \kappa \parallel \mathbf{r}_u \parallel^2} \tag{32}$$

where

$$\mathbf{r}_d = \mathbf{x}_d - \mathbf{c} \qquad \mathbf{r}_u = \mathbf{x}_u - \mathbf{c} \tag{33}$$

Here, $\mathbf{c}$ is the centre of radial distortion, $\mathbf{x}_u$ is the undistorted image point and $\mathbf{x}_d$ is the distorted image point. When the radial distortion is negative ($\kappa < 0$) images exhibit a barreling effect, whereas when distortion is positive ($\kappa > 0$) a pin-cushion effect appears in the image. Typically κ varies non-linearly with focal length and lies in the range $-1 \times 10^{-6} < \kappa < 1 \times 10^{-6}$. In Figure 12 we show the particular form of the distortion curve for the EIA servo-lens we have used in our experiments obtained using a calibration grid and a classical calibration algorithm.

The non-linear effect of radial distortion on the image formation process can be expressed as

$$\mathbf{x}_d = d(f, x_u)\ \mathtt{KRX} \tag{34}$$

where d is a non-linear function that depends on the focal length f and on the coordinates of the undistorted image point x_u. Tordoff and Murray [Tordoff and Murray2000] describe the geometric effect caused by radial distortion. This may be explained by the fact that rays are now bent and if we join a 3D point in space with its projection on the image by a straight line it will no longer pass through the centre of projection. This in turn implies that the transformation that maps corresponding points between views is no longer a homography and therefore the *infinite homography constraint* used for self-calibration is not strictly satisfied.

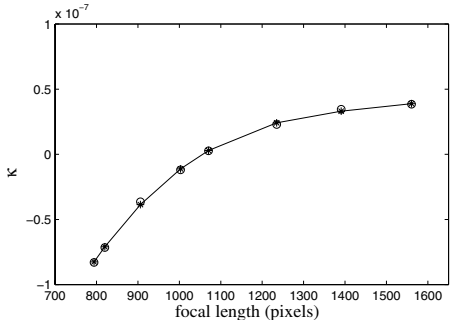


Fig. 12. Distortion curve for the EIA servolens used in our experiments showing the dependency of the distortion parameter κ with focal length

7.1. Effects of radial distortion on self-calibration

As one might expect from the fact that the infinite homography constraint is no longer satisfied, radial distortion has a severe impact on the self-calibration process. In this section we present some examples that demonstrate this situation.

In Figure 13 we show the ground truth values and the calibration parameters recovered by the non-linear self-calibration method on the same synthetic sequence used in previous experiments, but where the image points were first distorted according to the distortion curve shown in Figure 12. The focal length appears to be largely overestimated and the motion is no longer circular and is underestimated.

This same behaviour was observed on a real image sequence, where the camera performed a circular motion while zooming first in and then out. The images were not corrected for radial distortion. Figure 14 shows the self-calibration results obtained using the non-linear method imposing the square pixels and constant principal point constraints. Ground truth values for the motion and the calibration parameters (obtained using a classic calibration technique) are also shown for comparison.

Once more, the estimated motion is far from being circular and the focal length is underestimated. These results prove that it is crucial to introduce radial distortion into the self-calibration process, the question being how to do it automatically.

7.2. Solving for radial distortion

When the camera has constant focal length radial distortion can be described by a single κ. In that

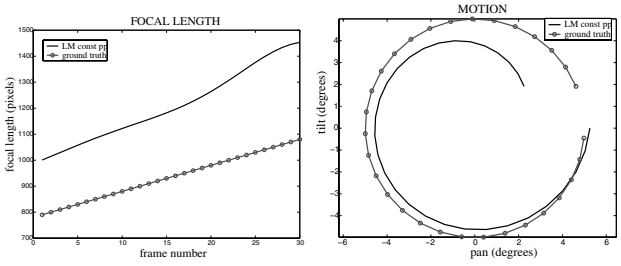


Fig. 13. Self-calibration results for a radially distorted synthetic image sequence. A raw self-calibration algorithm was used where radial distortion was not modelled.

case there are some well-behaved error measures that can be minimized to solve for the value of κ. Figure 15 depicts two such measures: the residual from the self-calibration minimization and the residual from the linear computation of the inter-image homographies from point matches. These cost functions give a good error surface and were used in previous experiments reported by Hartley [Hartley1997] and Sawhney *et al* [Sawhney and Kumar1999], where self-calibration and distortion were reliably determined for a non-zooming camera.

The main difficulty in solving for radial distortion when the focal length of the camera varies is that κ is different for each image. We have considered different ways to solve this problem.

- Include one κ_i parameter per image in the MLE minimization.
- Parameterize the distortion curve as a function of the focal length and solve for the curve parameters at the same time as the calibration using MLE.

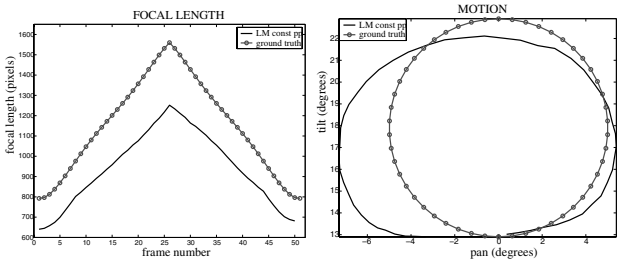


Fig. 14. Ground truth and self-calibration results for a radially distorted real image sequence.

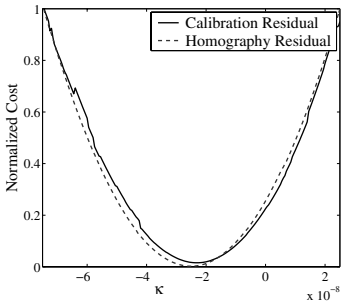


Fig. 15. Residual of the homography calculation and of the self-calibration algorithm for different values of *kappa* for a sequence with constant focal length.

- Correct the inter-image homographies for the radial distortion in an initial stage and then perform self-calibration using radially corrected data.

The first option can be quickly discarded. Experimentation shows that the large number of parameters causes the minimization to fall easily into local mimima.

The obvious disadvantage of the second option is that the parameterization would depend entirely on the specific dependence of the radial distortion parameter κ with the focal length for each lens. We have attempted to approximate the distortion curve by a 3^{rd} order polynomial which appears to describe well the behaviour of our EIA servolens (see Figure 12). We model the κ as

$$\kappa = \alpha_0 + \alpha_1 f + \alpha_2 f^2 + \alpha_3 f^3 \qquad (35)$$

and solve for $\alpha_0 ... \alpha_3$.

Figure 16 shows the results obtained for synthetic data using MLE to solve for the 4 distortion curve parameters at the same time as the calibration and motion parameters. Clearly, MLE fails to obtain good calibration results. The starting point for the minimization (LM constant pp) given by the output of the non-linear self-calibration method is too far from the solution, causing MLE to fall into a local minimum. The focal length and the motion parameters are improved but are still far from the true values.

Thus, the final solution appears to hold the most promise. It is based on the observation that an error in radial distortion correction will induce residual errors in the homography estimation in addition to noise.

While in the case of constant focal length this residual provided a good cost function for the com-

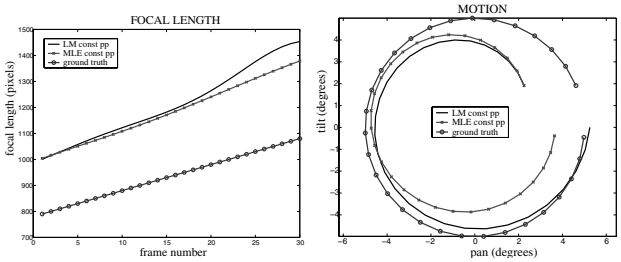


Fig. 16. Results for MLE including radial distortion terms in the minimization

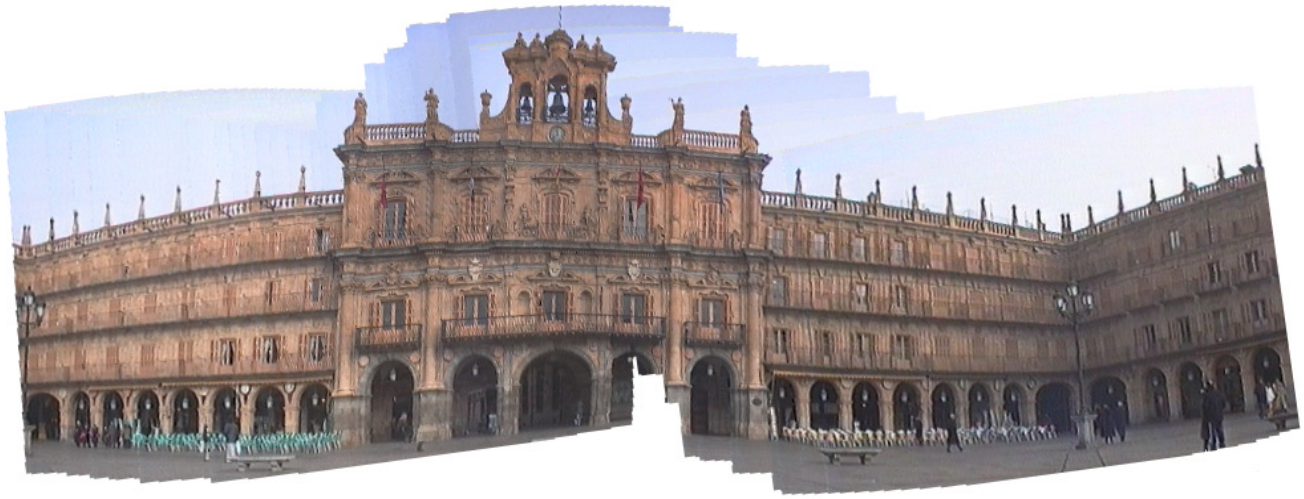


Fig. 17. Cylindrical mosaic created using 150 images of the Plaza Mayor in Salamanca. The camera's focal length varied throughout the sequence

putation of the radial distortion parameter, the situation is greatly complicated when we allow changing focal length (and hence κ). Tordoff and Murray [Tordoff and Murray2000] have developed a method that gives promising results on synthetic data. Here, cubic polynomials, describing locally the κ curve, are fitted to patches of 10 frames. Once the patches are all fitted, an average is taken where they overlap and this value of κ is used to perform the radial correction of each one of the images, and self-calibration would follow on the corrected images. This is the subject of ongoing investigation.

8. Application: automatic generation of cylindrical and spherical panoramic mosaics

Panoramic mosaics are generated by registering a series of images taken by a rotating camera and projecting them onto a cylinder or sphere. If the true focal length of the camera is known then the radius of the cylinder can be adjusted to that value. This would achieve perfect closure in the cylindrical projection in the case of a 360° panorama. If the focal length is not correctly estimated an error would occur in the length of the final composite image.

Most panoramic mosaicing techniques have been applied to image sequences taken with a camera with constant focal length [Szeliski and Shum1997, Sawhney et al.1998]. The difficulty with a sequence where the camera might be zooming between views is the accurate estimation of the focal length.

We have used our self-calibration method to calibrate a sequence of 150 images of the Plaza Mayor in Salamanca taken with a hand-held video camera where the camera was changing its zoom setting between views. The computed focal length was used to generate a cylindrical mosaic which is shown in Figure 17. Although the sequence only shows a 90° sweep this technique could be applied to generate full 360° panoramas.

9. Discussion

In this paper we have discussed the theory of self-calibration of rotating cameras with varying internal parameters. We have described two self-calibration algorithms (linear and non-linear) to solve for the varying intrinsic parameters of the camera given the inter-image homographies that relate corresponding points between views. The

algorithms are based on the use of the *infinite homography constraint* which describes the mapping of the image of the absolute conic (or its dual) between views. Both methods require some constraints to be imposed on the calibration parameters of the camera.

The first algorithm is iterative and requires an initial estimate, however we have experienced that the global minimum is achieved over a wide range of starting points. The problem is parameterized explicitly in terms of the calibration parameters of each camera, therefore it is a very flexible algorithm in terms of the constraints that can be imposed on the intrinsics. Each calibration parameter may be assumed to be known, constant throughout the sequence or free to vary. The linear algorithm is a fast and simple method, suitable for real-time applications, but is more restrictive in terms of the constraints that can be imposed on the internal parameters of the camera. It can be used with the minimal assumption of zero camera skew, but useful constraints such as a constant principal point or aspect ratio may not be imposed. A final bundle-adjustment algorithm where global image reprojection error is minimized has also been described. If some prior knowledge on the distribution of the parameters is known, this may be imposed via a MAP estimate.

In the experimental results we have analysed the effect of imposing different constraints on the intrinsic parameters of the camera. A relevant issue we have addressed is that, in general, best results are obtained when the principal point is assumed to be constant throughout the sequence, even when it is known to be varying in reality. This, perhaps contradictory, effect is caused by the fact that there is an inherent near-ambiguity in the simultaneous computation of the rotation parameters and the focal length of the camera which becomes more prominent when the principal point, a poorly conditioned parameter, is allowed to vary indiscriminately. Results improve when a MAP estimate is used to impose a prior on the location of the principal point.

An important issue we have addressed is the effect of radial distortion on the self-calibration process. Radial distortion can have very negative effects on the self-calibration of a rotating and zooming camera. Here we have investigated the possibility of including some radial correction parameters in the bundle-adjustment stage of the self-calibration process. However, in these experiments bundle-adjustment did not succeed to converge to the global minimum. Therefore, finding a reliable method to determine the radial distortion parameters automatically when the intrinsic parameters of the camera are varying is an important matter currently under investigation.

10. Acknowledgements

This work has been supported by the UK Engineering and Physical Science Research Council through Grant GR/L58668 and an Advanced Research Fellowship (I.D.R.), by the Norwegian Research Council (E.H.) and by the Spanish Ministry of Education and Science (L.A.).

References

[Agapito et al.1999] Agapito, L. de, R. I. Hartley, and E. Hayman: 1999, 'Linear calibration of a rotating and zooming camera'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado*. pp. 15–21.

[Agapito et al.1998] Agapito, L. de, E. Hayman, and I. Reid: 1998, 'Self-calibration of a rotating camera with varying intrinsic parameters'. In: *Proc. 9th British Machine Vision Conference, Southampton*. pp. 105–114.

[Agapito et al.2001] Agapito, L. de, I. Reid, E. Hayman, and R. Hartley: 2001, *Panoramic Vision*, Chapt. Self-calibration of zooming cameras from a single viewpoint: application to panoramic mosaicing. R. Benosman and S.B. Kang eds. Springer Verlag.

[Capel and Zisserman1998] Capel, D. and A. Zisserman: 1998, 'Automated Mosaicing with Super-resolution Zoom'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara*. pp. 885–891.

[Faugeras et al.1992] Faugeras, O. D., Q. Luong, and S. Maybank: 1992, 'Camera Self-Calibration: Theory and Experiments'. In: *Proc. European Conference on Computer Vision*. pp. 321–334, Springer-Verlag.

[Hartley1993] Hartley, R. I.: 1993, 'Euclidean reconstruction from uncalibrated views'. In: J. L. Mundy, A. Zisserman, and D. Forsyth (eds.): *Proc. 2nd European-US Workshop on Invariance, Azores*. pp. 187–202.

[Hartley1994] Hartley, R. I.: 1994, 'Self-Calibration from Multiple Views with a Rotating Camera'. In: *Proc. 3rd European Conference on Computer Vision, Stockholm*, Vol. 1. pp. 471–478.

[Hartley1997] Hartley, R. I.: 1997, 'Self-calibration of Stationary Cameras'. *International Journal of Computer Vision* **22**(1), 5–23.

[Hartley et al.1999] Hartley, R. I., E. Hayman, L. de Agapito, and I. D. Reid: 1999, 'Camera calibration and the search for infinity'. In: *Proc. 7th International*

Conference on Computer Vision, Kerkyra, Greece. pp. 510–517.

[Hayman2000] Hayman, E.: 2000, 'The use of zoom within active vision'. Ph.D. thesis, University of Oxford. Available at http://www.robots.ox.ac.uk/ActiveVision/Papers/... HaymanDPhil2000.

[Hayman et al.2000] Hayman, E., L. de Agapito, I. D. Reid, and D. W. Murray: 2000, 'The Rôle of Self-Calibration in Euclidean Reconstruction from Two Rotating and Zooming Cameras'. In: *Proc. 6th European Conference on Computer Vision, Dublin, Ireland*. pp. 477–492.

[Heyden and Åström1997] Heyden, A. and K. Åström: 1997, 'Euclidean reconstruction from image sequences with varying and unknown focal length and principal point'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. pp. 438–443.

[Heyden and Åström1998] Heyden, A. and K. Åström: 1998, 'Minimal Conditions on Intrinsic Parameters for Euclidean Reconstruction'. In: *Proc. Asian Conference on Computer Vision*.

[Heyden and Åström1999] Heyden, A. and K. Åström: 1999, 'Flexible Calibration: Minimal Cases for Auto-Calibration'. In: *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*. pp. 350–355.

[Heyden and Åström1996] Heyden, A. and K. Åström: 1996, 'Euclidean Reconstruction from Constant Intrinsic Parameters'. In: *Proc. International Conference on Pattern Recognition*.

[Kahl1999] Kahl, F.: 1999, 'Critical Motions and Ambiguous Euclidean Reconstructions in Auto-Calibration'. In: *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*. pp. 469–475.

[Kahl et al.2000] Kahl, F., B. Triggs, and K. Åström: 2000, 'Critical Motions for Auto-Calibration When Some Intrinsic Parameters Can Vary'. *Journal of Mathematical Imaging and Vision* **13**(2), 131–146.

[Liebowitz and Zisserman1998] Liebowitz, D. and A. Zisserman: 1998, 'Metric Rectification for Perspective Images of Planes'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. pp. 482–488.

[Luong and Viéville1996] Luong, Q. T. and T. Viéville: 1996, 'Canonical Representations for the Geometries of Multiple Projective Views'. *Computer Vision and Image Understanding* **64**(2), 193–229.

[Moons et al.1993] Moons, T., L. van Gool, M. van Diest, and A. Oosterlinck: 1993, 'Affine structure from perspective image pairs under relative translations between object and camera'. Technical Report KUL/ESAT/M12/9306, Departement Elektrotechniek, Katholieke Universiteit Leuven, Belgium.

[Pollefeys1999] Pollefeys, M.: 1999, 'Self-calibration and metric 3D reconstruction from uncalibrated image sequences'. Ph.D. thesis, ESAT-PSI, K.U.Leuven.

[Pollefeys et al.1998] Pollefeys, M., R. Koch, and L. Van Gool: 1998, 'Self Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters'. In: *Proc. 6th International Conference on Computer Vision, Bombay, India*. pp. 90–96.

[Sawhney et al.1998] Sawhney, H., S. Hsu, and R. Kumar: 1998, 'Robust Video Mosaicing Through Topology Inference and Local to Global Alignment'. In: *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*. pp. 103–119.

[Sawhney and Kumar1999] Sawhney, H. and R. Kumar: 1999, 'True Multi-Image Alignment and Its Application to Mosaicing and Lens Distortion Correction'. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(3), 245–243.

[Seo and Hong1998] Seo, Y. and K. Hong: 1998, 'Auto-calibration of a rotating and zooming camera'. In: *Proc. of IAPR workshop on Machine Vision Applications, Chiba, Japan*. pp. 17–19.

[Seo and Hong1999] Seo, Y. and K. Hong: 1999, 'About the Self-calibration of a rotating and zooming camera: Theory and Practice'. In: *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*. pp. 183–189.

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

[Sturm1997] Sturm, P.: 1997, 'Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1100–1105.

[Sturm1999] Sturm, P.: 1999, 'Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length'. In: *Proc. 10th British Machine Vision Conference, Nottingham*.

[Sturm and Maybank1999] Sturm, P. and S. Maybank: 1999, 'On plane based camera calibration: A general algorithm, singularities, applications'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition*. pp. 432–437.

[Szeliski and Shum1997] Szeliski, R. and H. Shum: 1997, 'Creating Full View Panoramic Image mosaics and Environment Maps'. In: *Proceedings of the ACM SIGGRAPH Conference on Computer Graphics*. pp. 251–258.

[Tordoff and Murray2000] Tordoff, B. J. and D. W. Murray: 2000, 'Violating Rotating Camera Geometry: The Effect of Radial Distortion on Self-Calibration'. In: *Proc. International Conference on Pattern Recognition*.

[Triggs1998] Triggs, B.: 1998, 'Autocalibration from Planar Scenes'. In: *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*.

[Triggs1997] Triggs, W.: 1997, 'Auto-Calibration and the Absolute Quadric'. In: *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico*. pp. 609–614.

[Tsai and Lenz1989] Tsai, R. and R. Lenz: 1989, 'A new technique for fully autonomous and efficient 3D robotics hand/eye calibration'. *IEEE Trans. Robotics and Automation* **5**(3), 345–358.

[Zhang1999] Zhang, Z.: 1999, 'A flexible new technique for camera calibration'. In: *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*.

[Zisserman et al.1998] Zisserman, A., D. Liebowitz, and M. Armstrong: 1998, 'Resolving Ambiguities in Auto-Calibration'. *Philosophical Transactions of the Royal Society of London, SERIES A* **356**(1740), 1193–1211.