The Rôle of Self-Calibration in Euclidean Reconstruction from Two Rotating and Zooming Cameras

Eric Hayman, Lourdes de Agapito, Ian D. Reid, and David W. Murray *

Department of Engineering Science, University of Oxford
Parks Road, Oxford OX1 3PJ, UK
{hayman,lourdes,ian,dwm}@robots.ox.ac.uk
http://www.robots.ox.ac.uk/ActiveVision/

Abstract. Reconstructing the scene from image sequences captured by moving cameras with varying intrinsic parameters is one of the major achievements of computer vision research in recent years. However, there remain gaps in the knowledge of what is reliably recoverable when the camera motion is constrained to move in particular ways. This paper considers the special case of multiple cameras whose optic centres are fixed in space, but which are allowed to rotate and zoom freely, an arrangement seen widely in practical applications. The analysis is restricted to two such cameras, although the methods are readily extended to more than two.

As a starting point an initial self-calibration of each camera is obtained independently. The first contribution of this paper is to provide an analysis of near-ambiguities which commonly arise in the self-calibration of rotating cameras. Secondly we demonstrate how their effects may be mitigated by exploiting the epipolar geometry. Results on simulated and real data are presented to demonstrate how a number of self-calibration methods perform, including a final bundle-adjustment of all motion and structure parameters.

1 Introduction

A configuration of cameras which occurs commonly in a number of imaging applications is that of multiple well-separated cameras whose optic centres are fixed in space, but which freely and independently (i) rotate about their optical centres and (ii) zoom in and out. This arrangement is used in surveillance and in broadcasting (particularly outside broadcasting), and is a pattern for acquiring models for virtual and augmented reality, where full or partial panoramas are taken from different positions around a building, for example.

What are the ways of handling the combined imagery from, say, two such uncalibrated cameras to recover a Euclidean reconstruction of a static scene? The least committed approach might be to generate a projective reconstruction, enforcing the zero translation

* This work is supported by the UK Engineering and Physical Science Research Council through Grant GR/L58668 (D.W.M) 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.).

D. Vernon (Ed.): ECCV 2000, LNCS 1843, pp. 477–492, 2000.

constraint within the images from a single camera, but then using overall a self-calibration algorithm for general motion, such as those of Pollefeys et al. [11], Heyden and Åström [9] or Hartley et al. [7]. A practical disadvantage is that general motion methods require locating the plane at infinity, but a broader criticism is that the motion is far from general.

The most committed (and perhaps most obvious) approach is to self-calibrate each rotating camera independently, for which methods have been described in the literature [3,2,12,13]. The task of reconstruction is then reduced to the more familiar one of structure from multiple views using calibrated cameras. Although dealing with each camera separately is attractive since it reduces the problem to a set of smaller, less complex ones, this method is likely to give poor results if the initial self-calibration is inaccurate. Moreover it is clearly not using all the available information.

The results in this paper provide two pieces of information which inform the solution from the spectrum of those available.

- There are some near-ambiguities in self-calibration of rotating cameras which can have a large effect both on the camera intrinsics and a reconstruction obtained from them. These effects are present when the self-calibration problem is ill-conditioned, in particular with small motions, large focal lengths, short image sequences and a poor spread of image features. They can be mitigated by modelling them correctly.
- Modelling the appropriate degree of inter-camera coupling is desirable. It proves useful to exploit the epipolar geometry not only to recover the relative positions of the two cameras, but also to refine the self-calibration of both sets of intrinsic parameters.

Two very different measures are used to characterize performance. The rms *reconstruction error* measures the distance between points in two rescaled and aligned Euclidean reconstructions. The rms *reprojection error* measures the faithfulness of reconstruction in the image: a low value implies that coplanarity and collinearity are well preserved, but it provides little information regarding the preservation of angles.

Results from any point in the spectrum of solutions may always be used to initialize a bundle-adjustment over all scene points and motion parameters, minimizing reprojection error. However, in addition to its cost implications for on-line use, bundle-adjustment is susceptible to convergence to local minima. The latter is critical in this context where a number of near-ambiguities are present since bundle-adjustment tends to make only very small changes to the motion parameters. Hence even if the reprojection error is reduced, there is by no means a guarantee of a significant change in reconstruction error. One goal of this work is to find algorithms which are good enough either to make bundle-adjustment unnecessary or to provide better initial estimates to increase the chance of convergence to the correct solution within it.

After introducing briefly the theory of self-calibration of rotating and zooming cameras in Section 2 we investigate precisely what information can and cannot be reliably extracted from such algorithms in Section 3. In particular we describe two near-ambiguities which commonly arise. In Section 4 we review the structure from motion algorithm of two calibrated views which we modify to resolve the ambiguities while minimizing epipolar transfer error. Experiments on synthetic and real data are presented in Section 5.

2 Self-Calibrating Rotating and Zooming Cameras: Review

The imaging process is modelled by the pinhole camera model so that in the ith image, the projection $\mathbf{x}_i$ of a point $\mathbf{X}$ in the scene is described by the relation $\mathbf{x}_i = \mathtt{P}_i\mathbf{X}$ where $\mathbf{x}_i$ and $\mathbf{X}$ are both given in homogeneous coordinates, implying that all vectors, matrices and equations are only defined up to an unknown scale factor. $\mathtt{P}_i$ is a 3×4 projection matrix which may be decomposed as $\mathtt{P}_i = \mathtt{K}_i\,(\mathtt{R}_i \;\; \mathbf{t}_i)$ where $\mathtt{R}_i$ and $\mathbf{t}_i$ describe the transformation between a coordinate frame attached to the scene and a camera centred coordinate system. $\mathtt{K}_i$ is the matrix of intrinsic parameters in image i and has the usual form

$$\mathtt{K} = \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad . \tag{1}$$

α_u and α_v are the focal lengths in the u and v directions, $(u_0\ v_0)$ are the coordinates of the principal point, and s is a parameter that describes the skew between the two axes of the CCD array.

In the case of a camera rotating about its optic centre, $\mathbf{t} = \mathbf{0}$, the final coordinate of $\mathbf{X} = (X\ Y\ Z\ 1)^\top$ is immaterial, and the projection equation simplifies to $\mathbf{x}_i = \mathtt{K}_i\mathtt{R}_i\,(X\ Y\ Z)^\top$. Different images taken from the same rotating camera relate to each other by homographies which take the form

$$\mathbf{x}_j = \mathtt{H}_{ij}\mathbf{x}_i = \mathtt{K}_j\mathtt{R}_j\mathtt{R}_i^{-1}\mathtt{K}_i^{-1}\mathbf{x}_i = \mathtt{K}_j\mathtt{R}_{ij}\mathtt{K}_i^{-1}\mathbf{x}_i \quad . \tag{2}$$

The inter-image homographies $\mathtt{H}_{ij}$ may be calculated directly from image measurements, for instance from point or line correspondences. Various techniques for this calculation are available, ranging from fast linear methods minimizing an algebraic error, via non-linear methods which minimize the geometric transfer error, to a bundle-adjustment in the motion and structure parameters, the structure comprises points on a mosaic.

Eliminating $\mathtt{R}_{ij}$ from equation (2) yields

$$\left(\mathtt{K}_j\mathtt{K}_j{}^\top\right) = \mathtt{H}_{ij}\left(\mathtt{K}_i\mathtt{K}_i{}^\top\right)\mathtt{H}_{ij}{}^\top \quad , \tag{3}$$

which can also be derived by projecting a point on the plane at infinity, $\mathbf{X} = (X\ Y\ Z\ 0)^\top$, into 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, and equation (3) is known as the *infinite homography constraint*. $\boldsymbol{\omega}^* = \mathtt{K}_i\mathtt{K}_i{}^\top$ is the dual of the image of the absolute conic (DIAC).

Given the homographies, $\mathtt{H}_{ij}$, equation (3) provides constraints on the intrinsic parameters. If the camera intrinsics are constant throughout the sequence, the constraint reduces to that of Hartley in [6], the DIAC may be computed linearly, and the matrix $\mathtt{K}$ is found from it by Cholesky decomposition.

For varying intrinsics, de Agapito et al. [3] solve equation (3) in a manner similar to that of Pollefeys et al. [11] for cameras undergoing general motion. In a non-linear

optimization the cost function

$$\mathcal{D} = \sum_{i=1}^{n} \| \mathtt{K}_i \mathtt{K}_i{}^{\top} - \mathtt{H}_{0i} \mathtt{K}_0 \mathtt{K}_0{}^{\top} \mathtt{H}_{0i}{}^{\top} \|_F^2 \tag{4}$$

is minimized, where the elements of $\mathtt{K}_i, i = 0...n$, are the unknown parameters. To eliminate the unknown scale factors, $\mathtt{K}_i \mathtt{K}_i{}^{\top}$ and $\mathtt{H}_{0i} \mathtt{K}_0 \mathtt{K}_0{}^{\top} \mathtt{H}_{0i}{}^{\top}$ are normalized so that their Frobenius norms are equal to one. An advantage of this approach is that any constraints on the intrinsic parameters, such as zero skew or known aspect ratio may be applied directly. Alternatively, parameters such as the aspect ratio or principal point can be solved for, but constrained to be constant throughout the sequence. A similar approach was adopted by Seo and Hong [12], but under known skew and principal point they note that the focal lengths can be computed linearly from equation (3).

In a later work [2] de Agapito et al. proposed a fast linear method for calculating all intrinsic parameters by employing an algebraic trick, used previously in another context by Armstrong et al. [1]. They dealt not with the DIAC, but with its *inverse*, the image of the absolute conic (IAC), $\boldsymbol{\omega}$. Under the assumption of zero skew the IAC is given by

$$\boldsymbol{\omega} = \mathtt{K}^{-\top} \mathtt{K}^{-1} = \begin{pmatrix} 1/\alpha_u^2 & 0 & -u_0/\alpha_u^2 \\ 0 & 1/\alpha_v^2 & -v_0/\alpha_v^2 \\ -u_0/\alpha_u^2 & -v_0/\alpha_v^2 & 1 + u_0^2/\alpha_u^2 + v_0^2/\alpha_v^2 \end{pmatrix} . \tag{5}$$

Inverting the infinite homography constraint, $\boldsymbol{\omega}_j = \mathtt{H}_{ij}{}^{-\top} \boldsymbol{\omega}_i \mathtt{H}_{ij}^{-1}$, provides linear constraints on the IAC in frame i by setting the (1,2) element of $\boldsymbol{\omega}_j$ to zero. Further constraints are available from additional assumptions on the intrinsic parameters, in particular, a known aspect ratio and/or a known principal point.

Most recently, optimal results have been obtained by de Agapito et al. [4] by performing a final bundle-adjustment in the motion and structure parameters.

2.1 Recovering Rotation Matrices and Euclidean Projection Matrices

For reconstruction, Euclidean projection matrices of the form $\mathtt{P}_i = (\overline{\mathtt{P}}_i \ \ \mathbf{0}) = (\mathtt{K}_i \mathtt{R}_i \ \ \mathbf{0})$ are required. The 3×3 left sub-matrices, $\overline{\mathtt{P}}_i$, are recovered from the projective homographies as $\mathtt{H}_{0i} \mathtt{K}_0$. Rotation matrices, referred to the initial frame, may be found by QR decomposition of $\overline{\mathtt{P}}_i$.

A more direct approach for finding rotations would be to use the recovered $\mathtt{K}_i$ matrices directly in the equation $\mathtt{R}_i = \mathtt{K}_i^{-1} \mathtt{H}_{0i} \mathtt{K}_0$. However, it can be unwise to apply this equation in combination with the non-linear self-calibration method of [3], especially when the principal point is constrained to be constant throughout the sequence in the minimization. The reason is that with fewer parameters in the model, the $\mathtt{R}_i$ recovered from $\mathtt{K}_i^{-1} \mathtt{H}_{0i} \mathtt{K}_0$ are less close to orthonormal. Even fitting an orthonormal matrix to $\mathtt{R}_i$ by setting the singular values of its SVD to unity does not guarantee that this rotation matrix is the correct one, especially since it is an algebraic error (a Frobenius norm) that is minimized when projecting $\mathtt{K}_i^{-1} \mathtt{H}_{0i} \mathtt{K}_0$ onto the 3-dimensional space of orthonormal matrices. This method could therefore give poor motion recovery and have dire consequences for Euclidean reconstruction. Rays would be back-projected incorrectly, and a large reprojection error

ensue. In this work we therefore adopt the approach based on QR decomposition when using the non-linear self-calibration algorithm.

Since the linear self-calibration method of [2] is not parameterized directly in terms of camera intrinsics, it does not suffer from the problems of non-orthonormal matrices, and the two approaches for recovering $\mathtt{R}_i$ are equivalent.

With pan-tilt cameras rotations are described by two rather than three parameters. A practical treatment of the decomposition of rotation matrices into these two parameters is provided in [8].

3 Ambiguities in Self-Calibration

Self-calibration is an ill-conditioned problem. Significant advances have been made since the work of Maybank and Faugeras, but there are a few underlying ambiguities which can have a large effect on results in configurations which poorly constrain the solution, coupled changes in the parameters in the model are barely observable. We consider two ambiguities present in the case of rotating cameras. It would be more correct to call these *near*-ambiguities: as opposed to true ambiguities which arise from certain motions and scenes [14,17], ours are only apparent because perspective effects in the cameras are less prominent under some camera configurations. A discussion of their relevance to reconstruction, motivated by experimental results, is provided in Section 6.

3.1 The Ambiguity between the Angle of Rotation and the Focal Length

For small rotations there is an 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. The ambiguity is easily seen by differentiating the calibrated non-homogeneous projection equation $\mathbf{x} = (\alpha/Z)\mathbf{X}$. Remembering that there is no translation, and secondly that the focal length α is a function of time, differentiation yields the following image motion in the x-direction

$$\dot{x} = +\alpha\Omega_Y - y\Omega_Z + \frac{x}{\alpha}\left(x\Omega_Y - y\Omega_X\right) + \frac{\dot{\alpha}x}{\alpha} \quad , \tag{6}$$

where $\mathbf{x}$ and $\mathbf{X}$ are expressed in camera centred frames and $\boldsymbol{\Omega}$ is the angular velocity.

Cyclorotation and the *relative* change in focal length can be recovered from the terms $-y\Omega_Z$ and $\dot{\alpha}x/\alpha$ respectively (the latter is zoom-induced looming motion). However, the first term $\alpha\Omega_Y$, a uniform motion in the image due to the component of rotation perpendicular to the optic axis, contains an ambiguity between focal length and rotation. The third term $(x/\alpha)(x\Omega_Y - y\Omega_X)$, which also arises from the component of rotation perpendicular to the optic axis, provides some disambiguating information, but the term is likely to be small except at the edges of the image. Unfortunately this is also where the optical properties of the lens are poorest. Notice too that the disambiguating information is weakest for large focal lengths. Compounding these difficulties is that in practical applications, sequences taken at large α are less likely to contain significant rotation. Since motion is being integrated over time, this ambiguity persists over a sequence of images.

In experiments we find that the 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. However, if the sequence is ill-conditioned, the ambiguity is also noticeable even if the principal point is constrained to a constant location.

3.2 The Principal Point/Rotation 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. If δu is the error in the estimation of the principal point, and α is the focal length, the erroneous rotation is $\sim \delta u/\alpha$ about y. This is an ambiguity between parameters from a single image.

Another way of describing this ambiguity is that a large focal length perspective projection is hard to distinguish from a spherical projection where the principal point is meaningless.

3.3 Experiments

Figures 1 and 2 illustrate these ambiguities using both simulated and real image data. The "bookshelf" sequence [3], was gathered by zooming while moving the vergence and elevation axes of one camera of a stereo head (equivalent to pan and tilt, up to an ordering of the the kinematic chain) so that the optic axis traced a right-circular cone. Point features were detected and matched, and homographies derived. Figure 1 shows the resulting mosaic. Simulated point data were synthesized similarly. Levenberg-Marquardt was used to minimize $\mathcal{D}$ in equation (4) allowing the principal point and focal length to vary over the sequence during minimization. (That is, in the minimization there are different values to be found for each frame, rather than a single value to be found for the whole sequence.)

In each set of results the first two plots show the recovered and veridical focal length and principal point. The $+$ symbols in the third plots show the recovered camera motion in terms of elevation and vergence angles. These are roughly circular, but there is a good deal of scatter about the best-fit circle.

In earlier work it was supposed that this scatter arose from noise [3]. However it turns out to be almost entirely due to the principal-point/rotation ambiguity. Using the ground truth value for the position of the principal point, the elevation and vergence angles are corrected and re-plotted as $\times$ symbols. These form near perfect circles.

However, the scale of the motion is still incorrect. This is due to the ambiguity between focal length/motion. 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.

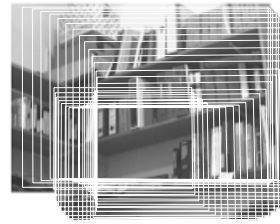

Fig. 1. A Mosaic constructed from the bookshelf sequence during which the camera panned and tilted while the lens zoomed.

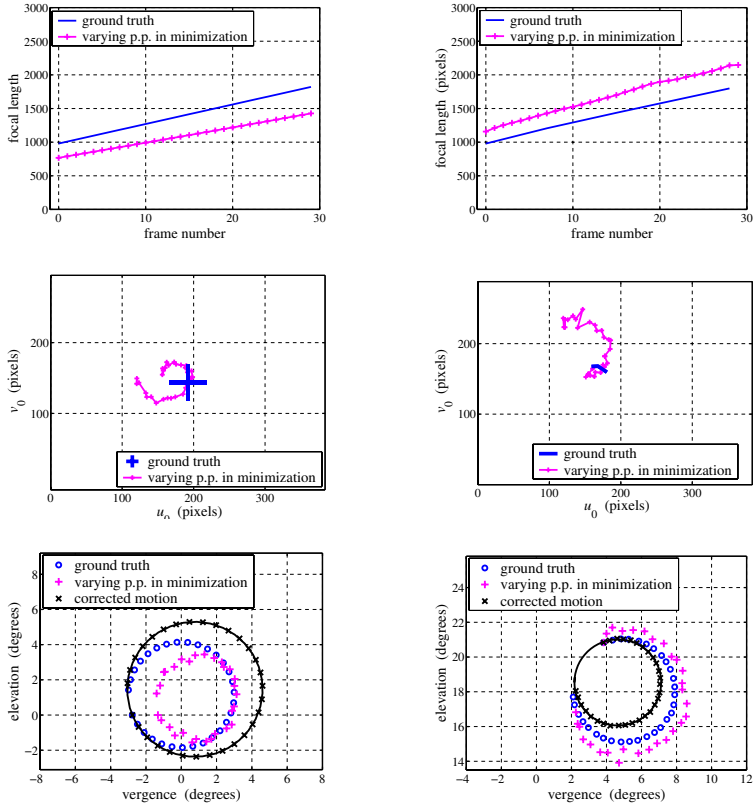

Fig. 2. Correcting the elevation and vergence angles by accepting the principal-point/rotation ambiguity. Parts (a) uses simulated data, (b) real data. Both sequences use a linearly increasing focal length and motion with cone half angle 3°.

Table 1. Verification of the ambiguity between focal length and motion exhibited in Figure 2.

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

4 Improving Self-Calibration via the Epipolar Geometry

We now turn to the second theme of the paper; using the appropriate degree of coupling between the two rotating cameras to improve the self-calibration of each, and then also improve the reconstruction. The methods utilize epipolar geometry and it is convenient first to review the work of, among others, Longuet-Higgins [10] and Zhang [16].

4.1 Stereo from Calibrated Cameras: Review

The geometry of two calibrated views is encapsulated in the essential matrix, $\mathtt{E}$ [10]. Corresponding image points $\mathbf{x}$ and $\mathbf{x}'$ in the first and second views are related by $\mathbf{x}'^{\top}\mathtt{E}\mathbf{x} = 0$ where $\mathtt{E} = [\mathbf{t}]_{\times}\mathtt{R}$, and $[\mathbf{t}]_{\times}$ is the skew-symmetric form of the translation vector, $\mathbf{t}$, describing the location of the optic centre of the second camera in the coordinate frame of the first. $\mathtt{R}$ is the relative rotation of the two cameras. $\mathtt{E}$ has five degrees of freedom, three for the rotation and two for translation up to scale. Since $[\mathbf{t}]_{\times}$ is rank two, so is $\mathtt{E}$, and the nullspace of $\mathtt{E}$ is $\mathbf{t}$. $\mathtt{R}$ may be recovered from $\mathtt{E}$ and $\mathbf{t}$ using quaternions.

The solution is refined with an algorithm due to Zhang [16] which uses the *uncalibrated* image measurements directly. For uncalibrated views the fundamental matrix, $\mathtt{F}$, plays a similar role to $\mathtt{E}$, and the two matrices relate as $\mathtt{F} = \mathtt{K}'^{-\top}\mathtt{E}\mathtt{K}^{-1}$ where $\mathtt{K}$ and $\mathtt{K}'$ are the intrinsic parameters for the first and second camera respectively. $\mathtt{F}$ is calculated directly from image measurements. An initial estimate of $\mathtt{F}$ is provided by the linear 8-point algorithm. The fundamental matrix is refined by minimizing a cost function with geometric significance, the distance between points and epipolar lines,

$$\mathcal{E} = \sum_k d^2(\mathbf{x}'^k, \mathtt{F}\mathbf{x}^k) + d^2(\mathbf{x}^k, \mathtt{F}^{\top}\mathbf{x}'^k), \quad d(\mathbf{x}^k, \mathtt{F}^{\top}\mathbf{x}'^k) = \frac{\mathbf{x}^{k\top}\mathtt{F}\mathbf{x}'^k}{\sqrt{(\mathtt{F}\mathbf{x}'^k)_1^2 + (\mathtt{F}\mathbf{x}'^k)_2^2}} \tag{7}$$

where the superscripts denote a particular point correspondence and $(\mathtt{F}\mathbf{x}'^k)_j$ is the jth component of the vector $(\mathtt{F}\mathbf{x}'^k)$. Thus, given $\mathtt{F}$ and the calibration matrices, $\mathtt{E}$ may be recovered and decomposed. The five parameters in $\mathtt{R}$ and $\mathbf{t}$ are then refined using the same geometric measure as above. $\mathbf{t}$ is parameterized by a point on the unit sphere, and $\mathtt{R}$ by a rotation vector[1].

Having computed $\mathtt{R}$, $\mathbf{t}$ and the self-calibration of each camera the scene may now readily be reconstructed, using not just the images from one stereo pair, but also further images. The projection matrices from the first and second cameras, $\mathtt{P}$ and $\mathtt{P}'$, in images i and i' respectively take the form

$$\mathtt{P}_i = \mathtt{K}_i\mathtt{R}_i\,(\mathtt{I} \quad \mathbf{0}) \qquad \text{and} \qquad \mathtt{P}'_{i'} = \mathtt{K}'_{i'}\mathtt{R}'_{i'}\,(\mathtt{R} \;\; \mathbf{t}) \quad . \tag{8}$$

3D points are found by the intersection of rays back-projected using these camera matrices. We will evaluate this algorithm in the experimental section.

[1] Zhang also performs a final bundle-adjustment over these five parameters and the 3D structure.

4.2 Constraints from the Epipolar Geometry of Two Rotating Cameras

The method given above uses only a single stereo pair to compute $\mathtt{R}$ and $\mathbf{t}$ and is clearly discarding a lot of information. Although we now have sufficient information to obtain a Euclidean reconstruction from the entire sequence, the result will be heavily biased towards the first pair.

Besides, since the fundamental matrix has seven degrees of freedom, and the essential matrix only five, it is possible to solve for two further parameters in $\mathtt{K}$ and $\mathtt{K}'$ just from the single pair. This is indeed done by Hartley in [5] and Pollefeys *et al.* in [11] who use linear methods to solve for the focal lengths assuming the principal points, aspect ratios and skew are known. However, in our case the special geometry may be used to greater effect by relating additional frames in either sequence to the original frame via the inter-image homographies.

We now write the epipolar constraint between correspondence k in image i from the first camera and image i' from the second as

$$\mathbf{x}'^{k\top}_{i'}\mathtt{F}_{ii'}\mathbf{x}^k_i = 0 \ . \tag{9}$$

As before, quantities without a dash refer to camera 1 and those with a dash to camera 2, subscripts relate to the frame number and superscripts to point correspondences. Choosing a reference frame from either camera gives

$$\mathtt{F}_{00} = \mathtt{K}'^{-\top}_0[\mathbf{t}]_\times \mathtt{R}\mathtt{K}_0^{-1} \ . \tag{10}$$

The fundamental matrix between two further images i and i' from each rotating camera relate to $\mathtt{F}_{00}$ as

$$\mathtt{F}_{ii'} = \mathtt{H}'^{-\top}_{i'}\mathtt{F}_{00}\mathtt{H}_i^{-1} \ . \tag{11}$$

Parameterizing $\mathtt{F}_{ii'}$ in terms of $\mathtt{F}_{00}$, points from several image pairs are used to refine our estimate of $\mathtt{F}_{00}$, and thus also $\mathtt{R}$ and $\mathbf{t}$. This is the second reconstruction algorithm we will investigate. The cost function minimized is the sum of epipolar distances over all measured points and also all images pairs,

$$\mathcal{F} = \sum_{i,i'}\sum_k d^2(\mathbf{x}'^k_{i'}, \mathtt{F}_{ii'}\mathbf{x}^k_i) + d^2(\mathbf{x}^k_i, \mathtt{F}_{ii'}{}^\top\mathbf{x}'^k_{i'}) \ . \tag{12}$$

Any combinations of i and i' may be chosen, provided image correspondences are available. Since $\mathcal{F}$ is a cost function with geometric significance there is a strong correlation with the reprojection error, but it is not the optimal error.

4.3 Improving Self-Calibration and Reconstruction

Now, since $\mathtt{H}_i = \mathtt{K}_i\mathtt{R}_i\mathtt{K}_0^{-1}$ we have that

$$\mathtt{F}_{ii'} = \mathtt{K}'^{-\top}_{i'}\mathtt{R}'_{i'}[\mathbf{t}]_\times \mathtt{R}\mathtt{R}_i{}^\top\mathtt{K}_i^{-1} \ . \tag{13}$$

Thus, estimates of (i) the relative camera positions $\mathtt{R}$ and $\mathbf{t}$ in a reference frame, (ii) the intrinsic parameters in both cameras at each frame in the sequence, and (iii) the

rotations between frames within sequences from either camera, yield an estimate of the fundamental matrix $\mathtt{F}_{ii'}$ between further frames of cameras 1 and 2. The goodness of this *modelled* fundamental matrix may then be measured with the cost function $\mathcal{F}$ in equation (12). Not only may further image pairs be used to provide further constraints on $\mathtt{R}$ and $\mathbf{t}$, further parameters may be solved for. Effectively we are constraining the inter-image homographies together with fundamental matrices. This insight provides the basis of the methods we derive for improving the self-calibration, and thus also the reconstruction.

We now introduce two methods of self-calibration refinement, depending on which ambiguities of Section 3 we wish to resolve. To parameterize the unknowns we write the *true* matrix of intrinsic parameters as

$$\mathtt{K}_i = \begin{pmatrix} \frac{1}{\beta}\hat{\alpha_i} & 0 & (u_0)_i \\ 0 & \frac{1}{\beta}\hat{\alpha_i} & (v_0)_i \\ 0 & 0 & 1 \end{pmatrix} \tag{14}$$

where $\hat{\alpha_i}$ is the *measured* focal length recovered from independent self-calibration of the rotating cameras, and where β is the unknown overall scale factor of the focal lengths of this camera over the entire sequence. Skew is assumed to be zero and the aspect ratio to be either known from the outset or recovered during self-calibration of each rotating camera. We also assume in both methods that the rotation matrices within a single camera have only two degrees of freedom, taking the form

$$\mathtt{R} = \mathtt{R}_y(\theta)\mathtt{R}_x(\phi) \ . \tag{15}$$

This is justifiable since pan-tilt cameras are restricted to this kind of motion (the ordering of $\mathtt{R}_y$ and $\mathtt{R}_x$ depends on the particular kinematic chain).

Method (1) deals only with the ambiguity between focal length and angle of rotation described in Section 3.1. Thus we solve for seven parameters, five for the motion and two for the overall scale of the focal lengths, β and β'. The true principal point (u_0, v_0) is assumed to be known from the self-calibration. The method is predicated on the assumption that rotations are small enough to model the ambiguity between focal length and rotation by requiring the *true* rotation matrix $\mathtt{R}_i$ to relate to the *measured* angles $\hat{\theta}_i$ and $\hat{\phi}_i$ as

$$\mathtt{R}_i = \mathtt{R}_y(\beta\hat{\theta}_i)\mathtt{R}_x(\beta\hat{\phi}_i) \ . \tag{16}$$

Method (2) seeks also to resolve the ambiguity between principal point and motion described in Section 3.2, and thus the number of parameters is $7 + 2n + 2n'$ where n and n' are the number of images from the two cameras. Method (2) models $\mathtt{R}_i$ by subtracting the erroneous motion caused by the ambiguity between rotations and motion of the principal point,

$$\mathtt{R}_i = \mathtt{R}_y\left(\beta\left(\hat{\theta}_i - \frac{u_0 - \hat{u_0}}{\hat{\alpha_i}}\right)\right)\mathtt{R}_x\left(\beta\left(\hat{\phi}_i - \frac{v_0 - \hat{v_0}}{\hat{\alpha_i}}\right)\right) \tag{17}$$

where $(\hat{u_0}, \hat{v_0})$ is the *measured* principal point from self-calibration of a rotating camera whereas (u_0, v_0) is its true value. The idea behind method (2) is based on the experimental

results of Section 3.3 where erroneous motion of the principal point is removed and the ambiguity between focal length and motion accounts for the remaining discrepancy from the ground truth.

4.4 Implementation Issues

Combining the information from two types of input, namely homographies and epipolar geometry, in order to provide accurate self-calibration and reconstruction places emphasis in methods (1) and (2) on retaining as much information from the initial self-calibration as possible. Two important issues are therefore *initialization* and *applying priors*.

In our current implementation an initial estimate of, β, and similarly β', are obtained by re-solving for only this single parameter in the non-linear self-calibration method. The prior is then found by investigating the curvature matrix $\mathtt{J}^\top \mathtt{J}$, where $\mathtt{J}$ is the Jacobian. In this case $\mathtt{J}^\top \mathtt{J}$ is a 1×1 matrix. In fact, experiments with a prior chosen more arbitrarily, and with β and β' initialized at unity, also worked well.

Furthermore, if the principal point was allowed to vary in the initial self-calibration, the correction devised in section 3.3 may be applied to initialize the principal point and motion in refinement methods (1) and (2). However, that example used ground truth of the principal point in the correction. Since such information is not available here, we initialize the principal point either at the centre of the image plane or with that obtained from the non-linear self-calibration method where the principal point is maintained at a fixed but unknown value throughout the sequence.

In our experiments we noticed that method (2) converges much more slowly than method (1). Therefore we choose only to use method (2) to refine the output from method (1).

4.5 Refining the Solution Using Bundle-Adjustment

The motion and structure parameters may be refined using a large non-linear minimization over all parameters, making use of the sparse form of the Jacobian. The cost function for the optimization is the reprojection error over all points and views,

$$\mathcal{C} = \sum_{\text{views}} \sum_{\text{points}} \|\mathbf{x} - \mathtt{K}\,(\mathtt{R}\ \ \mathbf{t})\,\mathbf{X}\|^2 \tag{18}$$

which provides a maximum likelihood estimate of the structure and motion. Each point $\mathbf{X}$ in the structure has either two or three degrees of freedom depending on whether it is visible from both cameras or only a single camera.

Bundle-adjustment is thus guaranteed to reduce the reprojection error, but not necessarily the reconstruction error. Of course the reconstruction gained is a valid Euclidean one in the sense that the projection matrices have the required form if parameterized as $\mathtt{P} = \mathtt{K}(\mathtt{R}\ \ \mathbf{t})$, but it may easily "look" more projective than Euclidean in that angles are skew, and length ratios are not preserved correctly. It would be naïve to expect bundle-adjustment to automatically cope with the inherent ambiguities which are present, the more so as it is prone to convergence to local rather than global minima. The parameters tend to change only by small amounts, and the final set of parameters differ little from the initial estimate.

5 Experiments and Results

Experiments were conducted first on simulated data to allow controlled investigation of the sensitivity of the reconstruction techniques to varying noise, and varying separation of the two cameras. The data were generated so as to correspond roughly with later experiments on real imagery. The image sizes were 384×288 pixels, and one camera had a focal length ranging from 1000 – 1870 pixels and a circular motion in the elevation and vergence axes of $4°$. The second camera had a longer focal length, 1250 – 2120 pixels, and a smaller circular motion of $3°$. The principal point used to generate the data moved between frames with an overall motion of approximately 20 pixels.

The self-calibration and reconstruction algorithms are summarized in Table 2.

Result 1. The principal result is that a significant improvement can indeed be achieved by our method of refining the self-calibration using epipolar geometry. In Figure 3 we compare both the reconstruction error and the reprojection error as a function of image position noise with no refinement of the self-calibration (using single and multiple views to calculate $\mathtt{R}$ and $\mathbf{t}$); and with refinement using method (1) of Section 4.3 which only handles the focal length/rotation ambiguity. Priors on the scale factor were obtained automatically from the method of Section 4.4.

Result 2. The performance of the algorithms with varying separation of the two cameras is shown in Figure 4. As before, a significant improvement may be obtained with our novel methods, especially when the cameras are close together, pointing in a similar average direction.

Whereas Result 1 used linear methods of recovering the homographies, and the linear method of initial self-calibration, and Levenberg-Marquardt for the minimization of the refinement cost function, and so is the fastest approach, this second experiment explores the other extreme. It uses bundle-adjusted homographies, the non-linear (LM) self-calibration, non-linear refinement, and finally bundle-adjusts the entire solution, solving for the focal length, principal point and two rotation parameters per camera, assuming square pixels. Furthermore, a longer image sequence (30 rather than 20 images) and more point correspondences (300 rather than 50 matches between images) were used. Again the refined method works much better than non-refined, and adding a final bundle-adjustment gives only a small further improvement. Notice that the results from method

Table 2. The algorithms evaluated in the experiments.

Algorithm outline	
Description	Label used in keys of graphs
A. Self-calibrate each camera individually	
B. Compute $\mathtt{R}$ and $\mathbf{t}$ from a single image pair	No refinement, single image pair
OR compute $\mathtt{R}$ and $\mathbf{t}$ from multiple image pairs	No refinement, multiple image pairs
C. Refine solution from B by resolving focal length/rotation ambiguity	Method (1)
D. (optional) Refine solution from C by resolving focal length/rotation and principal point/rotation ambiguities.	Method (2)
E. (optional) Bundle-adjustment, initialized at above solution	Bundle-adjustment

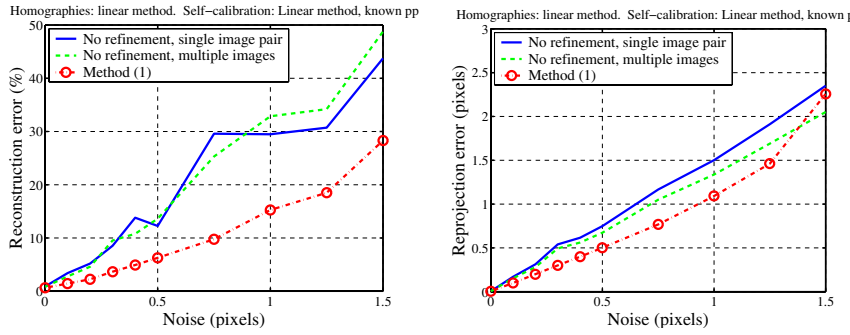

Fig. 3. The reconstruction and reprojection errors of different levels of noise, showing that the refinement of the self-calibration using epipolar geometry provides a significant improvement. (The angle between the principal directions of the cameras was fixed at 20°.)

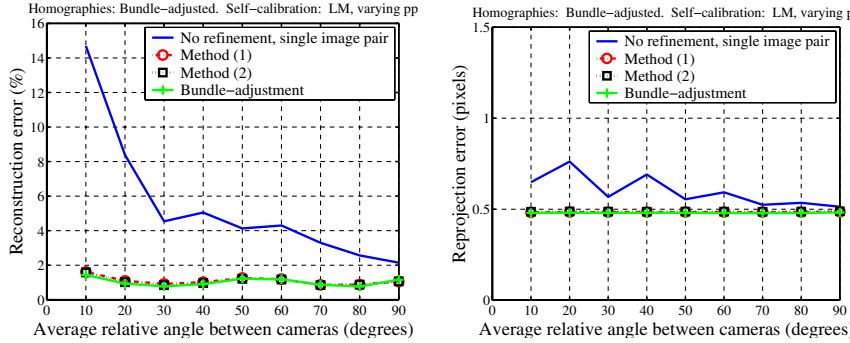

Fig. 4. The performance of the reconstruction techniques for different relative positions of the two cameras. Both cameras perform small rotational motions about some initial direction, the angle between the principal directions of the two cameras is plotted on the x-axis. In this experiment the noise was constant at $\sigma = 0.5$ pixels.

(2), which handles both ambiguities (focal length/rotation and principal point/rotation) is virtually indistinguishable from those from method (1) which handles only the former.

Result 3. Figure 5 demonstrates the sensitivity of bundle-adjustment to the initial estimate. A Euclidean bundle-adjustment is initialized with the output from the initial self-calibration, first with varying and then with fixed principal point in the minimization, and also with the output from our refinement method (2). Only small changes in parameters occur, and the reduction in reconstruction error is minimal.

5.1 Real Data

Two zoom sequences of a point grid, were taken with one of the cameras on a stereo head, using the common elevation and one of the vergence axes to generate the motion.

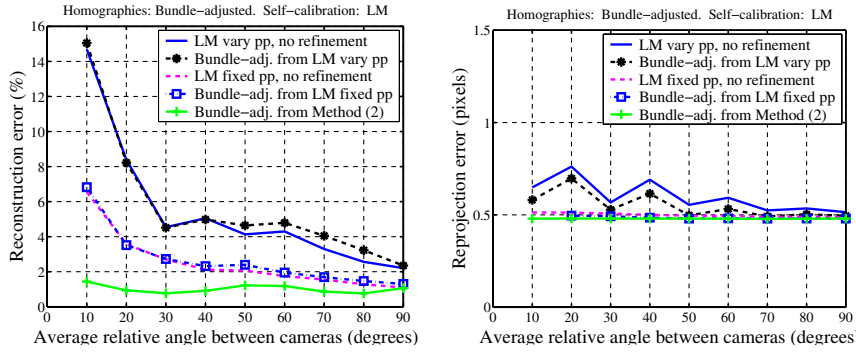

Fig. 5. A good initial estimate is crucial for bundle-adjustment. Using the present method for initialization yields much better results than when the bundle-adjustment is applied directly after self-calibration.

Fig. 6. The first, tenth and last images of the 20 frames of the sequences used from each camera. That the sequences were taken from viewpoints very close together is reflected in the similarity between the sequences.

Since we know the structure of the grid, we may measure results accurately, and the quality of the ensuing reconstruction is easily visualized. In the first sequence the focal length was varied between 1400 and 800 pixels (i.e. zooming out) with a circular motion of half-cone $2.5°$. In the second the focal length decreased from 1700 to 1100 pixels, and the circular motion was $2°$. Between the sequences the head was moved to provide a finite baseline. The angle between the scene and the two optic centres was approximately $10°$. The first, tenth and last images from a 20 image sequence are shown in Figure 6.

The motion in these sequences is very small, and the initial self-calibration was found to vary considerably depending on which algorithms were used to calculate the homographies and self-calibration, and how many images were used. Results from three experiments are summarized in Table 3, and Figure 7 shows reconstructions of the scene with and without refinement. Again, the novel methods presented in this paper provide a very significant improvement.

6 Conclusions

In this paper we have shown how systematic inaccuracies in the self-calibration of rotating cameras apparent in [3] can be accounted for by the ambiguities inherent in rotating motion fields. These effects are particularly keenly felt when small motions, large focal lengths, short image sequences and a poor spread of image features are involved.

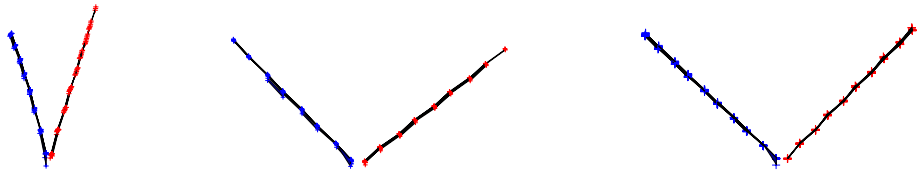

Fig. 7. Plan views of the reconstructed scene. (a) represents the fourth row of Table 3, using homographies calculated from the linear method, and the relative motion from a single image pair with no refinement. (b) demonstrates refinement method (1) applied to this reconstruction (row 6 in the table). (c) was obtained using bundle-adjusted homographies and a final bundle-adjustment of the motion and structure on the output from method (2) (row 10 in the table).

Table 3. Results of reconstruction of the calibration grid.

No. of images	Homography calculation	Self-calibration algorithm	Refinement method	Angle between planes	Reconstruction error (%)	Reprojection error (pixels)
5	Linear	Linear, known pp	Single image, no ref.	53.0	81.7	0.544
5	Linear	Linear, known pp	Multiple images, no ref.	52.7	79.0	0.489
5	Linear	Linear, known pp	Method (1)	109.7	19.9	0.298
8	Linear	Linear, known pp	Single image, no ref.	33.3	110.2	0.422
8	Linear	Linear, known pp	Multiple images, no ref.	93.4	8.3	0.316
8	Linear	Linear, known pp	Method (1)	95.7	8.0	0.259
20	Bundle-adj.	LM, varying pp	Single image, no ref.	98.8	9.2	0.336
20	Bundle-adj.	LM, varying pp	Method (1)	88.1	3.2	0.349
20	Bundle-adj.	LM, varying pp	Method (2)	88.8	3.0	0.361
20	Bundle-adj.	LM, varying pp	Bundle-adj.	90.6	2.8	0.327

The paper has also demonstrated that the epipolar geometry between multiple rotating cameras can and should be exploited to refine the initial self-calibration of the sets of intrinsic parameters, and hence to improve recovered scene structure. The improvements can be substantial.

By experiment, it has been shown too that, by itself, a Euclidean bundle-adjustment cannot resolve the ambiguities. Methods such as those presented here are required to initialize the adjustment. Interestingly, especially for those concerned with on-line time-sensitive implementations, the initialized position is often good enough for bundle-adjustment to make rather little improvement. In current work we are exploring the reduction of the parameters in bundle-adjustment just to those which appear poorly estimated from independent self-calibration of each camera. However it appears that the cost function surface of the reprojection error is still peppered with local minima.

The ambiguity between focal length and rotation is apparent as the bas-relief ambiguity in sequences with general motion, and can rapidly lead to disastrous results [15]. The reconstructed scene appears skewed relative to the true configuration, and length ratios are not preserved, implying that the upgrade from projective to Euclidean structure has not been successful. This is precisely the kind of behaviour we observe here (eg. Fig 7(a)) with reconstructions from multiple rotating cameras if the ambiguity between focal length and rotation is not resolved. It is found that the near ambiguity between

principal point and motion does not have as great an impact on the resulting Euclidean reconstruction, and resolving its effects are more difficult.

In future work we intend to extend the analysis to multi-focal constraints. This has the added benefit of better constrained matching.

References

[1] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In *Proc. European Conf. on Computer Vision*, LNCS 1064/5, pages 3–16. Springer-Verlag, 1996.

[2] L. de Agapito, R. I. Hartley, and E. Hayman. Linear calibration of a rotating and zooming camera. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado*, 1999.

[3] L. de Agapito, E. Hayman, and I. D. Reid. Self-calibration of a rotating camera with varying intrinsic parameters. In *Proc. BMVC*, pages 105–114, 1998.

[4] L. de Agapito, E. Hayman, and I. D. Reid. Self-calibration of rotating and zooming cameras. Technical Report TR 0225/00, University of Oxford, 2000.

[5] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. European Conf. on Computer Vision*, LNCS 588, pages 579–587. Springer-Verlag, 1992.

[6] R. I. Hartley. Self-calibration of stationary cameras. *International Journal of Computer Vision*, 22(1):5–23, February 1997.

[7] R. I. Hartley, E. Hayman, L. de Agapito, and I. D. Reid. Camera calibration and the search for infinity. In *Proc. 7th International Conference on Computer Vision, Kerkyra, Greece*, 1999.

[8] E. Hayman, J. G. Knight, and D. W. Murray. Self-alignment of an active head from observations of rotation matrices. To appear in *Proc. International Conference on Pattern Recognition*, 2000.

[9] A. Heyden and K. Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 1997.

[10] H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[11] M. Pollefeys, R. Koch, and L. Van Gool. Self calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proc. 6th Int'l Conf. on Computer Vision, Bombay*, pages 90–96, 1998.

[12] Y. Seo and K. Hong. Auto-calibration of a Rotating and Zooming Camera. In *Proc. of IAPR workshop on Machine Vision Applications*, pages 17–19, November 1998.

[13] Y. Seo and K. Hong. About the self-calibration of a rotating and zooming camera: Theory and practice. In *Proc. International Conference on Computer Vision*, 1999.

[14] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico*, pages 1100–1105, June 1997.

[15] R. Szeliski and S. B. Kang. Shape ambiguities in structure from motion. In *Proc. 4th European Conf. on Computer Vision, Cambridge*, pages 709–721, 1996.

[16] Z. Zhang. A new multistage approach to motion and structure estimation: From essential parameters to euclidean motion via fundamental matrix. Technical Report TR2910, INRIA Sophia-Antipolis, June 1996.

[17] A. Zisserman, D. Liebowitz, and M. Armstrong. Resolving ambiguities in auto-calibration. *Philosophical Transactions of the Royal Society of London, SERIES A*, 356(1740):1193–1211, 1998.