A Rational Function Lens Distortion Model for General Cameras

David Claus and Andrew W. Fitzgibbon
Department of Engineering Science
University of Oxford, Oxford OX1 3BN
{dclaus,awf}@robots.ox.ac.uk

Abstract

We introduce a new rational function (RF) model for radial lens distortion in wide-angle and catadioptric lenses, which allows the simultaneous linear estimation of motion and lens geometry from two uncalibrated views of a 3D scene. In contrast to existing models which admit such linear estimates, the new model is not specialized to any particular lens geometry, but is sufficiently general to model a variety of extreme distortions.

The key step is to define the mapping between image (pixel) coordinates and 3D rays in camera coordinates as a linear combination of nonlinear functions of the image coordinates. Like a "kernel trick", this allows a linear algorithm to estimate nonlinear models, and in particular offers a simple solution to the estimation of nonlinear image distortion. The model also yields an explicit form for the epipolar curves, allowing correspondence search to be efficiently guided by the epipolar geometry. We show results of an implementation of the RF model in estimating the geometry of a real camera lens from uncalibrated footage, and compare the estimate to one obtained using a calibration grid.

1. Introduction

A considerable volume of recent research has underlined the benefits of wide-angle and omnidirectional cameras in computer vision tasks. With a wide field of view, problems such as egomotion estimation, structure and motion recovery, and autocalibration are better conditioned, yielding more accurate answers with fewer images than are required with narrow-field of view perspective cameras. Until recently, however, such cameras have required careful assembly and precalibration before they can be used for 3D reconstruction. With the advent of variable focal-length wide-angle cameras and lenses, such precalibration is very difficult to achieve. In addition, many practical applications involve archive footage which has been captured by an un-
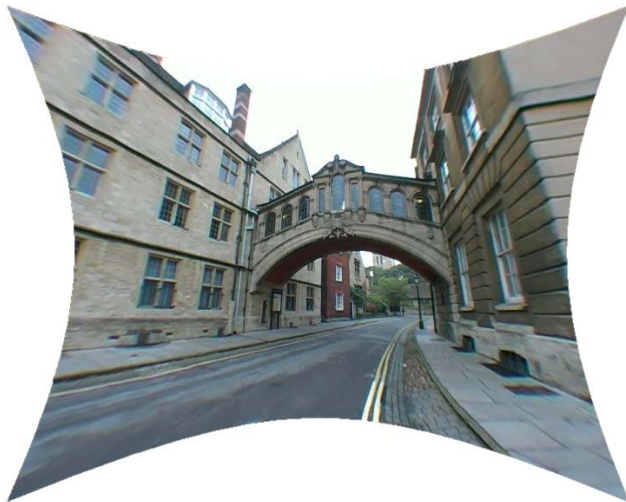

Figure 1. An image with lens distortion corrected using the Rational Function model described in this paper. Distortion can be computed from either a calibration target or from two or more images of the scene.

known lens, and for which even qualitative geometry is not available.

Recent advances [4, 5, 12] have shown how to extend the epipolar geometry of uncalibrated perspective cameras [8] to the non-pinhole case. However, the extensions have been limited to low distortion [4] or have been restricted to specific camera geometries [5, 12]. In this paper we show that a quite general model may be fit linearly to 2D image correspondences, recovering the distortion parameters and the scene structure up to a perspective transformation.

2. State of the Art

Wide angle images can be obtained from fish-eye lenses or catadioptric cameras. Both methods introduce a highly nonlinear mapping between image coordinates and 3D ray directions from the camera; it is the task of camera calibration to determine this mapping. In this paper we are particularly interested in the problem of calibrating central,

non-pinhole cameras using only 2D point correspondences obtained from a moving camera.

The generalization of epipolar geometry from pinhole cameras to catadioptric systems was first mooted by Svoboda [17], who defined epipolar curves for a family of cameras, but assumed a precalibrated system. Kang [9] showed how the system parameters and motion could be simultaneously extracted by nonlinear estimation, in a manner similar to that of Zhang's [18] technique for estimation of radial lens distortion. Fitzgibbon [4] showed that for low distortion, it is possible to obtain a solution for both epipolar geometry and a single parameter of lens distortion linearly from 2D point correspondences. Mičušík [12] extended this model to cope with a variety of wide-angle lenses and catadioptric cameras, but a general model is not yet known.

The closest work to that which we report here is that of Geyer and Daniilidis [5]. For parabolic catadioptric cameras, they describe an algorithm which permits linear estimation of egomotion and lens geometry. By lifting image points $\mathbf{x} = [i, \ j, \ 1]^\top$ to a four dimensional 'circle space',

$$\Pi(\mathbf{x}) = \begin{pmatrix} i \\ j \\ i^2 + j^2 - 1/4 \\ 1 \end{pmatrix} \quad (1)$$

they obtain a 4×4 analogue of the fundamental matrix, which can be fit to image points. From point correspondences, they can compute the camera calibration (focal length and image center) and epipolar geometry from two or three views.

Sturm [15] uses another lifting strategy to present models for back-projection of rays into images from affine, perspective, and para-catadioptric cameras. These models are then used to obtain fundamental matrices between multiple views and differing camera types, and epipolar geometry computed using these methods is illustrated.

2.1. Contribution of this paper

In this paper we extend the lifting strategies [5, 15] to build a general-purpose model for catadioptric cameras. Although the new model is algebraic in inspiration rather than geometric, we believe it is sufficiently powerful to model a range of highly distorted cameras. The model is a derivate of the rational cubic camera described by Hartley and Saxena [7], and here we show how it admits linear solutions for epipolar geometry estimation.

The paper is in three main parts. In §3 we discuss the new rational function model, showing examples of its modelling power, how to project and backproject ("distort" and "undistort") points, and how to fit it to calibration-grid data. Then §4 characterizes its epipolar geometry via a 6×6 analogue of the fundamental matrix, which we call G. It is
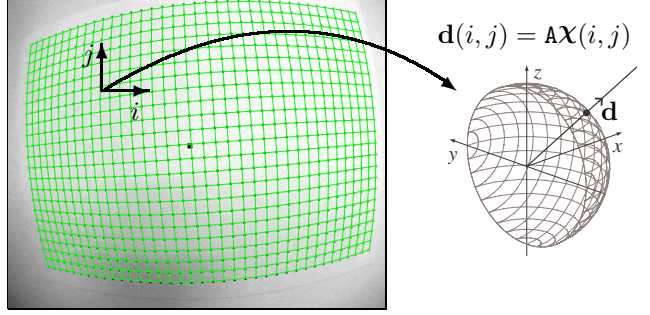

Figure 2. The task of distortion correction is to determine the mapping between image coordinates (i, j) and 3D rays $\mathbf{d}_{ij}$. The green lines superimposed on the image of the planar grid of calibration dots indicate the accuracy of the RF model.

shown how to fit it to point correspondences, and how to recover the distortion parameters from a fitted G. Finally we discuss in §5 the practical estimation of G, investigate its noise sensitivity, and demonstrate the successful recovery of epipolar curves on real image pairs.

3. Rational Function Model

As described in [13, 16], calibration of a camera amounts to determining the 3D ray from which each pixel samples. For a central camera, this means finding the mapping from pixels (i, j) in $\mathbb{R}^2$ to rays $\mathbf{d}(i, j)$ in $\mathbb{R}^3$. For a perspective camera, this mapping is a vector of linear polynomials in (i, j), viz

$$\mathbf{d}(i, j) = \begin{pmatrix} A_{11}i + A_{12}j + A_{13}1 \\ A_{21}i + A_{22}j + A_{23}1 \\ A_{31}i + A_{32}j + A_{33}1 \end{pmatrix} = \mathbf{A} \begin{pmatrix} i \\ j \\ 1 \end{pmatrix}, \quad (2)$$

where the 3×3 matrix $\mathbf{A} = \mathbf{KR}$ is the standard camera calibration matrix [8], typically upper triangular, but which could be full if camera coordinates are arbitrarily rotated. In this paper, we explore the behaviour of a mapping which permits i and j to appear in higher order polynomials, in particular quadratic:

$$\mathbf{d}(i, j) = \begin{pmatrix} A_{11}i^2 + A_{12}ij + A_{13}j^2 + A_{14}i + A_{15}j + A_{16} \\ A_{21}i^2 + A_{22}ij + A_{23}j^2 + A_{24}i + A_{25}j + A_{26} \\ A_{31}i^2 + A_{32}ij + A_{33}j^2 + A_{34}i + A_{35}j + A_{36} \end{pmatrix}. \quad (3)$$

This model may be written as a linear combination of the distortion parameters, in a 3×6 matrix A, and a 6-vector $\boldsymbol{\mathcal{X}}$ of monomials in i and j. Define $\boldsymbol{\mathcal{X}}$ as the "lifting" of image point (i, j) to a six dimensional space

$$\boldsymbol{\mathcal{X}}(i, j) = [i^2, \ ij, \ j^2, \ i, \ j, \ 1]^\top \quad (4)$$

Our imaging model is then given by

$$\mathbf{d}(i, j) = \mathbf{A}\boldsymbol{\mathcal{X}}(i, j) \quad (5)$$

$$
\begin{array}{ll}
\text{Rational Function} & (p,q) = \left(\dfrac{A_1^\top \boldsymbol{X}(i,j)}{A_3^\top \boldsymbol{X}(i,j)}, \dfrac{A_2^\top \boldsymbol{X}(i,j)}{A_3^\top \boldsymbol{X}(i,j)} \right) \\[3ex]
\text{Field of View [2]} & (p,q) = \dfrac{\tan(r_d \omega)}{2 r_d \tan(\omega/2)}(i_d, j_d) \\[3ex]
4^{th}\text{ order radial [8]} & (p,q) = (1 + k_1 r_d^2 + k_2 r_d^4)(i_i, j_d) \\[3ex]
\text{Division model [4]} & (p,q) = \dfrac{1}{(1 + k r_d^2)}(i_i, j_d)
\end{array}
$$

Bicubic model [10]
$$
\begin{aligned}
p &= b_1 i_d^3 + b_2 i_d^2 j_d + b_3 i_d j_d^2 + b_4 j_d^3 \\
 &\quad + b_5 i_d^2 + b_6 i_d j_d + b_7 j_d^2 + b_8 i_d + b_9 j_d + b_{10} \\
q &= c_1 i_d^3 + c_2 i_d^2 j_d + c_3 i_d j_d^2 + c_4 j_d^3 \\
 &\quad + c_5 i_d^2 + c_6 i_d j_d + c_7 j_d^2 + c_8 i_d + c_9 j_d + c_{10}
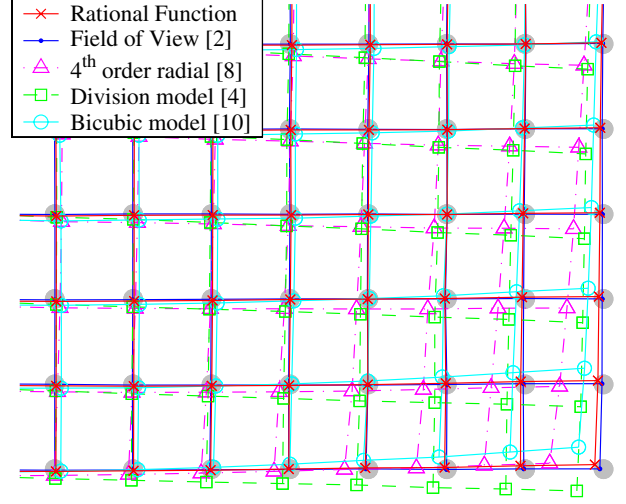\end{aligned}
$$

Figure 3. Comparison of distortion removal for the lower right corner of Figure 2. For each of the comparison models, distortion parameters and a planar homography were fit via nonlinear optimization. The RF model is comparable with the fish-eye specific FOV model, but permits linear estimation.

where $\mathbf{d}$ is a vector in camera coordinates representing the ray direction along which pixel (i, j) samples (see Figure 2).

Undistorted image coordinates (p, q) are computed by the perspective projection of $\mathbf{d}$:

$$
(p,q) = \left(\frac{A_1^\top \boldsymbol{X}(i,j)}{A_3^\top \boldsymbol{X}(i,j)}, \frac{A_2^\top \boldsymbol{X}(i,j)}{A_3^\top \boldsymbol{X}(i,j)} \right) \tag{6}
$$

where the rows of $\mathtt{A}$ are denoted by $A_{1..3}^\top$. We see that the mapping $(i, j) \rightarrow (p, q)$ is a quotient of polynomials in the image coordinates, hence the name *rational function* model [1].

3.1. Distortion model comparison

The mapping from image pixels (i, j) to distortion corrected points (p, q) can be viewed as a nonlinear distortion correction followed by a projective homography. We compared the distortion correction using several common models as detailed in Figure 3. Each model was fit via nonlinear optimization parameterized by: the pixel aspect ratio a, distortion center (i_c, j_c), the model parameters, and a planar homography. The error measure was the distances between the 850 mapped image points and their true grid locations. The distorted radius is given by $r_d = \sqrt{(i - i_c)^2 + a^2(j - j_c)^2}$.

The Field of View (FOV) model [2] is based on the fish-eye lens design parameters; this physical basis results in an excellent distortion fit. The flexibility of the RF model yields a very close approximation to both the FOV solution and the true grid, and can be fit linearly from either a calibration grid or two uncalibrated views. Note that the higher

order terms in the bicubic model [10] do not compensate for the lack of a rational function denominator.

3.2. Back-projection and projection

Back-projecting an image point to a ray in 3D space is done by simply lifting the point and applying (5). To *project* a 3D point $\mathbf{X}$ (in camera coordinates) into the image plane we must determine the (i, j) corresponding to the ray $\mathbf{d} = -\mathbf{X}$. As (5) is defined only up to scale, we must solve

$$
\mathbf{d} = \lambda \mathtt{A} \boldsymbol{X}(i,j) \tag{7}
$$

for i, j, and λ. Note that each row $A_{1..3}^\top$ of $\mathtt{A}$ may be identified with a conic in the image plane $A_i^\top \boldsymbol{X}(i,j) = 0$. If $\mathbf{d} = (d_1, d_2, d_3)^\top$, then cross-multiply to eliminate λ, so that (i, j) are the points of intersection of the pair of conics

$$
\begin{aligned}
(d_1 A_3 - d_3 A_1)^\top \boldsymbol{X}(i,j) &= 0 \\
(d_2 A_3 - d_3 A_2)^\top \boldsymbol{X}(i,j) &= 0
\end{aligned} \tag{8}
$$

yielding up to four possible image locations (i, j). In practice there are only two intersections; one corresponds to the pixel which images ray $\mathbf{d}$, and the other to $-\mathbf{d}$. For cameras with viewing angles less than $180°$ the correct intersection is closest to the center of the viewable area (the boundary of which is given by the conic A_3). More general cameras require a further check of the viewing quadrant of the point in order to determine the correct intersection.

3.3. Calibration grids

We now present a linear method for calibrating a camera using a single image of a planar calibration grid. The

grid (a sample employing circular dots is shown in Figure 2) is used to provide some number N of correspondences between (lifted) 2D points $\boldsymbol{\mathcal{X}}_k$ and 2D plane points $\mathbf{u}_k$. As the plane is in an unknown position and orientation, the mapping between plane and camera coordinates is a 3×3 homography $\mathtt{H}$. Thus each correspondence provides a constraint of the form

$$
\begin{align}
\mathtt{H}\mathbf{u}_k &= \lambda_k \mathtt{A}\boldsymbol{\mathcal{X}}_k \tag{9}\\
\implies \mathbf{u}_k &= \lambda_k \mathtt{H}^{-1}\mathtt{A}\boldsymbol{\mathcal{X}}_k \tag{10}\\
\implies \mathbf{u}_k &= \lambda_k \mathtt{A}'\boldsymbol{\mathcal{X}}_k \tag{11}\\
\implies \mathbf{0} &= \left[\mathbf{u}_k\right]_\times \mathtt{A}'\boldsymbol{\mathcal{X}}_k \tag{12}
\end{align}
$$

from which a linear solution for $\mathtt{A}'$ is obtained. Despite the presence of the unknown homography in $\mathtt{A}'$, it will still rectify the images to pinhole, so that lines are imaged as straight. Figure 1 illustrates an image corrected using such an $\mathtt{A}'$. Section 3.4 discusses the choice of homography in order to obtain a "canonical" $\mathtt{A}$.

3.4. Canonicalization of $\mathtt{A}$

In the above technique, the matrix of camera intrinsics and distortion coefficients $\mathtt{A}$ is recovered only up to an unknown homography on the "output side" of $\mathtt{A}$. Thus, if the camera's true intrinsics are $\mathtt{A}_{\text{true}}$, then the $\mathtt{A}$ recovered is related to $\mathtt{A}_{\text{true}}$ by an unknown projective homography

$$\mathtt{A}_{\text{true}} = \mathtt{H}\mathtt{A}. \tag{13}$$

This homography is analogous to the unknown homography encountered when recovering pinhole projection matrices $\mathtt{P}$ from a pinhole fundamental matrix. In some applications the choice of $\mathtt{H}$ is unimportant. For example, if image rectification were our goal, knowledge of $\mathtt{A}$ up to an unknown homography is all that is required, because the image corners should be mapped to the corners of the output image in order to best use the available image resolution.

In general, however, we would like to remove the ambiguity in $\mathtt{A}$ as much as possible, and this section discusses some of the constraints which we may bring to bear. We will do this by showing that $\mathtt{A}$ has a canonical form, to which an arbitrary $\mathtt{A}$ can be transformed, thus reducing the ambiguity.

First, we note that because the unknown homography appears on the "output side" of $\mathtt{A}$, it corresponds to an affine change of basis in $\mathbb{R}^3$. We are at liberty when calibrating our camera to choose this basis to some extent. For example, we may rotate camera coordinates arbitrarily and still have a calibrated camera, so we can without loss of generality insist that the camera Z-axis passes through the pixel $(i, j) = (0, 0)$. This implies that $\mathtt{A}\boldsymbol{\mathcal{X}}(0,0) \propto [0, 0, 1]^\top$, and thus, as $\boldsymbol{\mathcal{X}}(0,0) = [0, 0, 0, \ 0, 0, 1]^\top$, that the last column of $\mathtt{A}$ can be fixed to be of the form $[0, 0, \cdot]^\top$, where the dot denotes an arbitrary scalar.
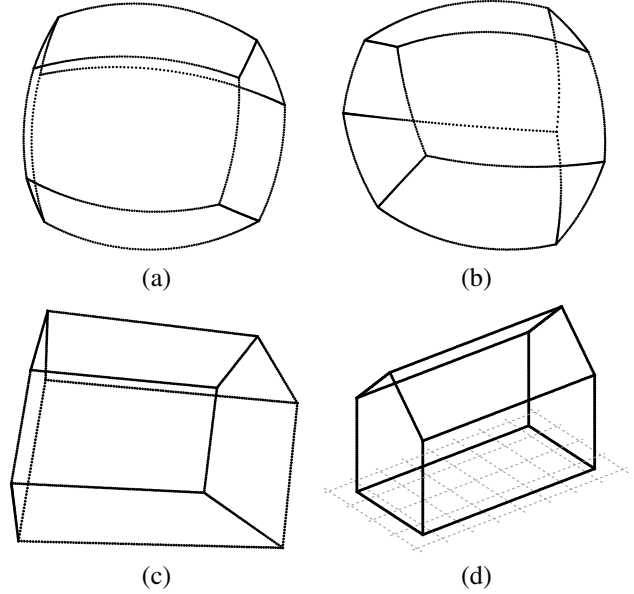

Figure 4. Rectification can be performed from two views and no additional information. (a),(b) two synthetic images of a house model (c) rectified structure computed from two views (d) original 3D model.

Note also that the conic described by the first row of $\mathtt{A}$ is the set of image points whose world ray's $d_1 = 0$; that is, the conic A_1 is the image of the world YZ plane. Similarly, conic A_2 is the image of the world XZ plane. We can further fix camera coordinates by requiring that the tangents to these two conics at the origin are aligned with the i and j axes. The normal to the conic $ai^2 + bij + cj^2 + di + ej + f = 0$ is $(2ai + bj + d, bi + 2cj + e)$, and evaluated at the origin, the normal is (d, e). In terms of $\mathtt{A}$ we require that the normal to A_1 is along $(1, 0)$ and that the normal of A_2 is along $(0, 1)$, fixing A_{14}, A_{15}, A_{24}, and A_{25}. Thus we have the canonical form of $\mathtt{A}$:

$$
\mathtt{A} = \begin{bmatrix} \times & \times & \times & \times & 0 & 0 \\ \times & \times & \times & 0 & \times & 0 \\ \times & \times & \times & \times & \times & \times \end{bmatrix}. \tag{14}
$$

From any of the algorithms in this paper which produce an $\mathtt{A}$ defined only up to a homography, we may always choose a 3D rotation which transforms to the above canonical representation. Thus the ambiguity is reduced to a four-parameter family characterized by an upper triangular matrix, which we shall call $\mathtt{K}$. Determination of $\mathtt{K}$ is a problem analogous to pinhole-camera self-calibration, and is the subject of ongoing investigation. In some simple cases it can be determined, for example the conic A_3 is the image of the world XY plane. Although the conic itself may be difficult to observe, its center can be seen in the images of some catadioptric systems, for example a camera viewing a spherical mirror, and thus used to further constrain $\mathtt{A}$.

Figure 5. Epipolar curves recovered from the G matrix computed using 36 point correspondences. Note that the curves converge to the camera viewpoints on either side of the street.

4. Two-view geometry

Without a calibration grid, we still would like to compute A. This section examines how that can be accomplished from two distorted images of a scene. A 3D point imaged in two views gives rise to a pair of point correspondences $(i, j) \leftrightarrow (i', j')$ in the image. Their back-projected rays $\mathbf{d}$, $\mathbf{d}'$ must intersect at the 3D point, that is to say they must satisfy the Essential matrix constraint

$$\mathbf{d}^\top \mathtt{E} \mathbf{d}' = 0. \tag{15}$$

Substituting $\mathbf{d} = \mathtt{A}\boldsymbol{\mathcal{X}}$, we obtain

$$\boldsymbol{\mathcal{X}}^\top \mathtt{A}^\top \mathtt{E} \mathtt{A}' \boldsymbol{\mathcal{X}}' = 0 \tag{16}$$

and writing $\mathtt{G} = \mathtt{A}^\top \mathtt{E} \mathtt{A}'$ we obtain the constraint satisfied by the lifted 2D correspondences

$$\boldsymbol{\mathcal{X}}^\top \mathtt{G} \boldsymbol{\mathcal{X}}' = 0. \tag{17}$$

This "lifted fundamental matrix" is rank two and can be computed from 36 point correspondences in a manner analogous to the 8-point algorithm [11] for projective views. We shall refer to this as the DLT. Although this can be done in a strictly linear manner, greater accuracy is achieved through a nonlinear optimization using the Sampson distance approximation [8], as discussed in §5. The immediate consequence is that, because G can be determined solely from image-plane correspondences, we can remove distortion from image pairs for which no calibration is available. Before describing how this is achieved, however, we first investigate the epipolar curves induced by the new model.

4.1. Epipolar curves

For a given pixel (i', j') in one of the images, its epipolar curve in the other image is easily computed, as the set of (i, j) for which $\boldsymbol{\mathcal{X}}(i, j)^\top \mathtt{G} \boldsymbol{\mathcal{X}}' = 0$. This is the conic with parameters $\mathtt{G}\boldsymbol{\mathcal{X}}'$. Figure 5 shows epipolar curves computed from point correspondences on two real-world images. We note that all the curves intersect at two points, because G is rank 2, so all possible epipolar conics are in the pencil of conics generated by any two columns of G. The two intersection points in each image are the two intersections of the baseline with the viewing sphere of that image.

4.2. Recovery of A from G

We now demonstrate that the distortion matrix A can be recovered from the matrix G, up to a projective homography. This A can then be used to rectify the two images of the scene. Once rectified, the images and 2D correspondences can be passed through standard structure from motion algorithms for multiple views (based on the pinhole projection model, and also only determined up to a projective ambiguity) to recover 3D scene geometry. Autocalibration techniques on the pinhole images can also upgrade A. We assume we are dealing with a moving camera with fixed internal parameters, so that $\mathtt{A}' = \mathtt{A}$.

Given G, we wish to decompose it into factors $\mathtt{A}^\top \mathtt{F} \mathtt{A}$, where F is rank two. It is clear that any such factorization can be replaced by an equivalent factorization $(\mathtt{H}\mathtt{A})^\top (\mathtt{H}^{-\top} \mathtt{F} \mathtt{H}^{-1})(\mathtt{H}\mathtt{A})$ so we can at best expect to recover A up to a premultiplying homography. By ensuring F has the form of an essential matrix this ambiguity can be reduced to a rotation of camera coordinates.

Our strategy for computing A is to extract its (3D) or-

thogonal complement from the 4D nullspace of G. We note that as F is rank 2, it has a right nullvector, which we shall call **e**. The nullspace of G is the set of X for which

$$\mathtt{A}^\top \mathtt{F}\mathtt{A}X = 0 \tag{18}$$

and because A has full rank, this is also the set of X for which

$$\mathtt{F}\mathtt{A}X = 0 \tag{19}$$

writing $\mathtt{N} = \mathrm{null}(\mathtt{G})$, a 6×4 matrix, we have that all such X are of the form $\mathtt{N}\mathbf{u}$ for $\mathbf{u} \in \mathbb{R}^4$, so

$$\mathtt{F}\mathtt{A}\mathtt{N}\mathbf{u} = 0 \ \ \forall \mathbf{u} \in \mathbb{R}^4 \tag{20}$$

so $\mathtt{A}\mathtt{N}\mathbf{u} \propto \mathbf{e} \ \forall \mathbf{u}$, i.e. all columns of $\mathtt{A}\mathtt{N}$ are multiples of **e**, so

$$\mathtt{A}\mathtt{N} = \mathbf{e}\mathbf{v}^\top \tag{21}$$

for some $\mathbf{v} \in \mathbb{R}^4$, and so

$$\mathtt{A} = \mathbf{e}\mathbf{v}^\top \mathtt{N}^\top + \mathtt{M}\mathtt{N}^\perp \tag{22}$$

where $\mathtt{N}^\perp = \mathrm{null}(\mathtt{N}^\top)^\top$ and M is an arbitrary 6×2 matrix. Choosing $\mathbf{e} = -\mathrm{null}((\mathtt{N}^\perp)_{[1:6,4:6]})$ and $\mathtt{M} = \mathrm{null}(\mathbf{e}^\top)$ means A is now a function only of $\mathbf{v}$. The 4×1 vector $\mathbf{v}$ is found by nonlinear optimization of the following residuals on the original points (where $\mathtt{A}^* = \mathtt{A}(\mathbf{v})$):

$$\sum \left[(\mathtt{A}^* \boldsymbol{\chi}'_k)^\top \mathtt{A}^* \mathtt{G}\mathtt{A}^{*\top} \mathtt{A}^* \boldsymbol{\chi}_k \right]^2. \tag{23}$$

Figure 4 shows the result of an implementation of this process on synthetic data generated using the real-world A obtained from our fisheye lens. The recovered A does indeed rectify the house, showing that the algorithm works in principle; the following section examines its noise sensitivity.

5. Practicalities

We present results on two experimental questions: how stable is the computation of G as measured by the quality of epipolar curves; and how effective is rectification using the recovered A.

Synthetic tests of the recovery of G were carried out (see Figure 6) with Gaussian noise added to the house images of Figure 4. The RMS value of the Sampson distances [14] from the points to their epipolar conics was used as the error measure. By removing the distortion using a calibrated A we can estimate a pinhole F and reconstruct $\mathtt{G} = \mathtt{A}^\top \mathtt{F}\mathtt{A}$. This reconstructed G provides a lower bound on the Sampson error. A nonlinear optimization initialized with the true G and evaluated on noisy image points converged to the same solution. This indicates that a stable minimum exists in the Sampson distance space, even at high noise levels.

Careful data conditioning is important for computing a fundamental matrix from noisy image correspondences [6].
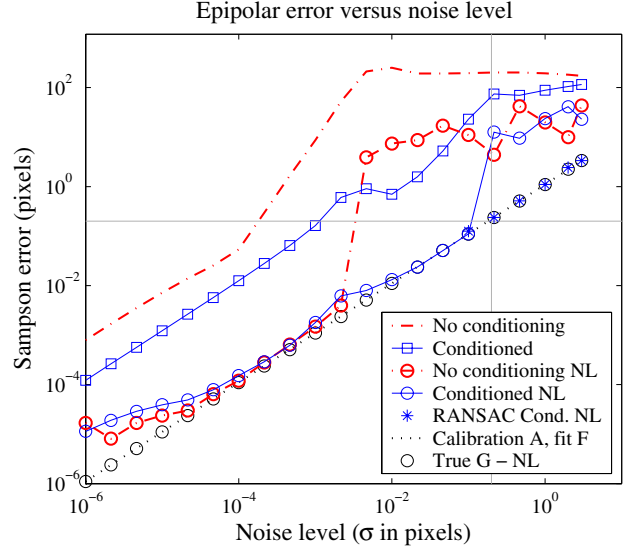

Figure 6. Sampson error for various noise levels applied to synthetic data. Removing distortion with a calibration A and then fitting F yields errors equal to the input noise. Auto-calibration is more sensitive to noise, but careful conditioning and nonlinear optimization (NL) provide better results. With RANSAC, the global Sampson minimum is found.

We conditioned the data by translation and scaling so the image points lie within $[-0.5 \ldots 0.5]$, then lifted the points before applying a scaling and translation so each dimension of the 6D vector had zero mean and RMS magnitude $\sqrt{2}$. The linear results were refined using a nonlinear optimization (MATLAB's lsqnonlin) with G parameterized as a sum of two rank-one matrices thus:

$$G = u_1 u_2^\top + u_3 u_4^\top, \ \ u_{1..4} \in \mathbb{R}^6. \tag{24}$$

At an amplitude of 10^{-3} pixels the noise begins to dominate the solution. Data conditioning is therefore essential, and G should be constructed from the DLT eigenvector which produces the smallest Sampson error. A RANSAC strategy should be employed for noise above 0.1 pixels (see Figure 6).

A digital camcorder equipped with a wide-angle lens was used to capture some real-world footage, and point correspondences were identified between Harris corners in each of two frames. In all, 135 correspondences were available to compute G. A linear fit followed by nonlinear optimization produced poor results, however, using a RANSAC strategy to generate initial estimates for a nonlinear fit produced a G with lower residuals on the correspondences (RMS Sampson error of 0.6 pixels), which generated the epipolar geometry shown in Figure 5.

The algorithm to recover the distortion matrix A is less successful. Good results are achieved up to a Sampson error of approximately 0.2 pixels, however beyond that the
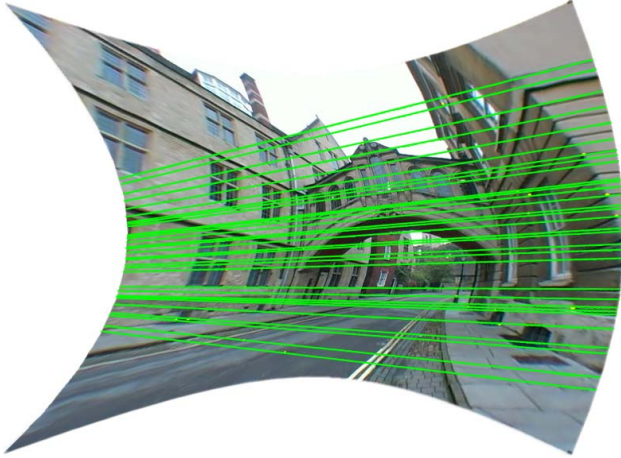
Figure 7. The left image from Figure 5 after rectification using an A recovered from G. Although the epipolar curves have correctly been straightened, the vertical lines are not straight. Augmenting the estimation with a plumbline constraint may correct this for sequences with such predominantly horizontal epipolar lines.

recovered A from the real example does not match that from the calibration grid, and produces inaccurate rectifications (see Figure 7). The following section will discuss our plans to improve this situation.

6. Conclusions

We have introduced a new model for wide-angle lenses which is general, and extremely simple. Its simplicity means it is easy to estimate both from calibration grids and from 2D-to-2D correspondences. The quality of the recovered epipolar geometry is encouraging, particularly given the large number of parameters in the model. It appears that the stability advantages conferred by a wide field of view act to strongly regularize the estimation.

However, although G appears well estimated, the algorithm to recover the distortion matrix A remains unsatisfactory. Our current work is looking at three main strategies for its improvement. First, the current algorithm is essentially only a linear algorithm given G, and does not minimize meaningful image quantities using a parametrization of the full model. Second, by combining multiple views with a common lens, we hope to obtain many more constraints on A and thus achieve reliable results. Third, we currently nowhere include the plumb-line constraint [3]. Augmenting the estimation with a single linearity constraint is the focus of our current work.

To put these results in perspective, an analogy with early work on autocalibration in pinhole-camera structure and motion recovery might be in order. Initial attempts focused on the theory of autocalibration, and showed its potential, but it was several years before practical algorithms were available. It is our claim that the simplicity and performance of the RF model shown in this paper commends it to the field as a practical general-purpose model for large field-of-view imaging systems.

Acknowledgements

References

[1] C. Clapham. *The Concise Oxford Dictionary of Mathematics*. Oxford University Press, second edition, 1996.

[2] F. Devernay and O. Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 13:14–24, 2001.

[3] F. Devernay and O. D. Faugeras. Automatic calibration and removal of distortion from scenes of structured environments. In *SPIE*, volume 2567, San Diego, CA, 1995.

[4] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. CVPR*, 2001.

[5] C. Geyer and K. Daniilidis. Structure and motion from uncalibrated catadioptric views. In *Proc. CVPR*, 2001.

[6] R. I. Hartley. In defense of the eight-point algorithm. *IEEE PAMI*, 19(6):580–593, 1997.

[7] R. I. Hartley and T. Saxena. The cubic rational polynomial camera model. In *Proc. DARPA Image Understanding Workshop*, pages 649–653, 1997.

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision, 2nd Ed.* Cambridge: CUP, 2003.

[9] S. B. Kang. Catadioptric self-calibration. In *Proc. CVPR*, volume 1, pages 201–207, 2000.

[10] E. Kilpelä. Compensation of systematic errors of image and model coordinates. *International Archives of Photogrammetry*, XXIII(B9):407–427, 1980.

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

[12] B. Mičušík and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. In *Proc. CVPR*, volume 1, pages 485–490, 2003.

[13] R. Pless. Using many cameras as one. In *Proc. CVPR*, volume 2, pages 587–593, 2003.

[14] P. D. Sampson. Fitting conic sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm. *CVGIP*, 18:97–108, 1982.

[15] P. Sturm. Mixing catadioptric and perspective cameras. In *Proc. Workshop on Omnidirectional Vision*, 2002.

[16] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *Proc. ECCV*, volume 2, pages 1–13, 2004.

[17] T. Svoboda, T. Pajdla, and T. Hlaváč. Epipolar geometry for panoramic cameras. In *Proc. ECCV*, pages 218–232, 1998.

[18] Z. Zhang. On the epipolar geometry between two images with lens distortion. In *Proc. ICPR*, pages 407–411, 1996.