Simultaneous linear estimation of multiple view geometry and lens distortion

Andrew W Fitzgibbon
Department of Engineering Science
The University of Oxford
19 Parks Road, Oxford, United Kingdom, OX1 3PJ

Abstract

A bugbear of uncalibrated stereo reconstruction is that cameras which deviate from the pinhole model have to be pre-calibrated in order to correct for nonlinear lens distortion. If they are not, and point correspondence is attempted using the uncorrected images, the matching constraints provided by the fundamental matrix must be set so loose that point matching is significantly hampered.

This paper shows how linear estimation of the fundamental matrix from two-view point correspondences may be augmented to include one term of radial lens distortion. This is achieved by (1) changing from the standard radial-lens model to another which (as we show) has equivalent power, but which takes a simpler form in homogeneous coordinates, and (2) expressing fundamental matrix estimation as a Quadratic Eigenvalue Problem (QEP), for which efficient algorithms are well known.

I derive the new estimator, and compare its performance against bundle-adjusted calibration-grid data. The new estimator is fast enough to be included in a RANSAC-based matching loop, and we show cases of matching being rendered possible by its use. I show how the same lens can be calibrated in a natural scene where the lack of straight lines precludes most previous techniques. The modification when the multi-view relation is a planar homography or trifocal tensor is described.

1. Introduction

This paper deals with the problem of nonlinear lens distortion in the context of camera self-calibration and structure from motion. In particular, the recovery of 3D camera motion from 2D point tracks, where there is moderate to severe radial lens distortion. The paper uses an unusual model for distortion which—like the most common existing model—is a reasonable approximation to the distortion observed in typical cameras. The advantage of the alternative approximation is an extension of uncalibrated projective reconstruction to the case of unknown lens distortion, thus closing an important gap in the applicability of uncalibrated vision. While the fundamental matrix allows a direct solution to the simultaneous recovery of relative orientation and camera calibration from a stereo pair, this paper's model allows a direct solution for the aforementioned parameters *and* a single lens distortion term. Figure 2 shows compellingly how adding a single distortion term can improve the results of scene reconstruction over long video sequences.

Figure 1: An image with lens distortion corrected using the technique described in this paper. Distortion can be computed given two or more images of the scene, and no other information.

The largest body of work on the estimation of lens distortion deals with precalibration, where the camera is calibrated *offline*. In this case, the 3D positions of scene points are known, and the direct linear transform [13, §16.2.3.2.1] allows recovery of the camera projection matrix and a single lens distortion parameter via a generalized eigensystem. We are interested in the situation where the original lens is not available, for example with archive footage, or when using variable lens geometries.

Previous work on the *online* estimation of lens distortion divides neatly into two strategic approaches. The first, known as the *plumb line* method [2, 3, 10, 14, 16] uses

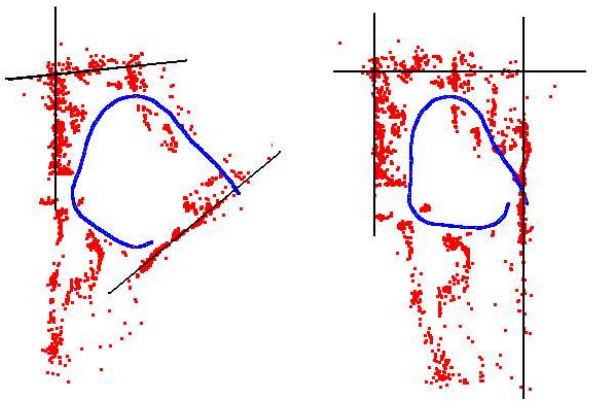


Figure 2: A 3D reconstruction of an 800-frame video of an office scene computed by a commercial camera tracker [1], without distortion correction (left) and with (right). The lines of the scene walls have been marked in to aid visualization. The uncorrected images yield a reconstruction in which focal length and rotation are incorrectly estimated, and which is therefore far from the correct geometry, while the result from corrected images is very close to orthogonality. A single lens distortion term was used, with the distortion center at the center of the image.

straight lines in the scene to provide constraints on the distortion parameters. However, straight lines are not always available in the scene (see figure 1 for example), and when present are not trivial to detect. Therefore, such methods often require care and supervision to ensure that real-world curves are not confused with distorted lines.

This paper is in the second class of methods, which demand nothing more of the images than the rigidity assumptions which allow the computation of the fundamental matrix. Such methods are exemplified by the work of Zhang [20] and Stein [15], where the rigidity constraint is extended to include the parameters of the distortion model. Until now, however, these and related techniques [4, 11, 12] have relied on iterative methods to find the distortion parameters. As is usual with such iterative methods, their convergence is not guaranteed, initial estimates must be found, and—although fast within the class of nonlinear routines—they remain too slow to place in the inner loop of any hypothesize-and-test architecture. The algorithm herein is a one-shot method relying on the solution of a generalized eigenproblem, whose convergence is well studied, and for which fast algorithms exist. In addition, the analysis of the new algorithm indicates that at least some formulations of the problem may have many local minima, making a good initialization even more necessary.

1.1. The goals of this paper

The primary goal of this paper is to allow the matching of image pairs via interest-point correspondences, especially when lens distortion would otherwise hinder the process. The most successful current techniques for matching interest points are based on the geometric constraints offered by multiple-view geometry [8]. These are effective because fast linear algorithms exist for the computation of the relationships, allowing their computation to form the kernel of RANSAC-based matching algorithms. However, when images have strong lens distortion, these constraints cannot be applied, because the two-view relationships (fundamental matrix, planar homography) are not valid in the image periphery.

Thus we seek to find a model for the between-view relations which incorporates lens distortion. In particular, we seek a model which admits a direct solution, i.e. computation via well understood, fast, and globally convergent, numerical algorithms such as the SVD or eigenvalue extraction.

Our goal in this paper is *not* the accurate estimation of the lens distortion coefficients themselves. If accurate camera information is required, there is no recourse but to bundle adjustment [18], initialized with (1) reasonable estimates of camera geometry and (2) good correspondences. It is in the provision of these two requisites that this paper contributes.

2. Notation

We shall denote 2D points (in non-homogeneous coordinates) by $\underline{\mathbf{x}} = (x, y)$ and let $\mathbf{x}$ denote a general vector, including 2D points in homogeneous coordinates. Matrices are represented in fixed-width font F. The data used by the new algorithm comprises point correspondences between lens-distorted images. As we shall deal almost entirely with two-view geometry, we shall use primes to indicate a corresponding point in the second view. Thus, as input we have a set of *two-view point correspondences*, denoted $\mathbf{x} \leftrightarrow \mathbf{x}'$.

The image points which we observe will be distorted functions of some perfect, pure, perspective, pinhole points, which we shall always denote using $\mathbf{p}$, so the image point $\underline{\mathbf{x}}$ is the distorted version of perfect point $\mathbf{p}$. This paper discusses only radial distortion, so that the relationship between $\underline{\mathbf{x}}$ and $\mathbf{p}$ is dependent on their distances from the image center. Throughout the paper, all these points are expressed in a 2D coordinate system with origin at the *distortion* center. The implications of this are discussed in §8.1.

Given, then, that the distortion center may be assumed known, the distortion model may be written

$$\underline{\mathbf{x}} \quad = \quad \mathcal{L}(\mathbf{p})$$

Indeed, when dealing with *radial* distortion, the mapping

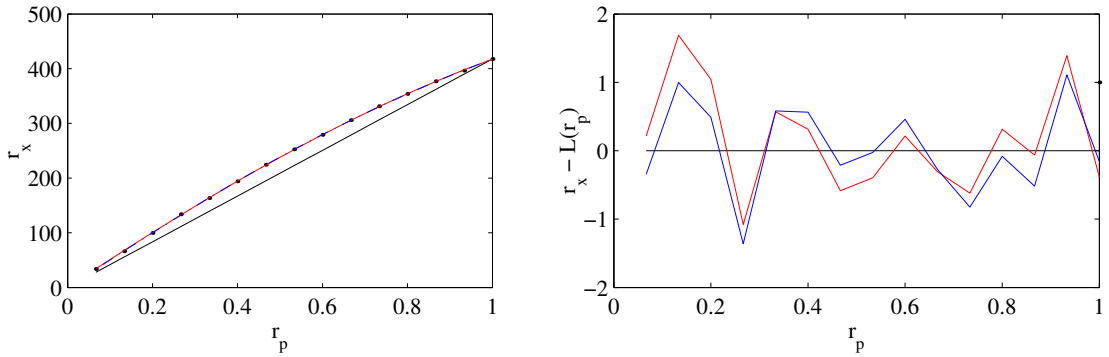


Figure 3: Comparison of lens distortion models with ground truth. The camera was fitted with a 4.2mm lens. *Left:* The Y axis shows the image radius $\|\mathbf{x}\|$, the X axis is the pinhole radius $\|\mathbf{p}\|$. Black dots mark the ground truth distortion curve, measured by imaging a fronto-parallel calibration grid. The red curve shows the best fit of the traditional model (equation 1), and the blue curve is the best fit to the division model (equation 2). The black line is $\|\mathbf{x}\| = \|\mathbf{p}\|$, for visualization. *Right:* Approximation errors $\|\mathbf{x}\| - L(\|\mathbf{p}\|)$ for the two models.

is simply between the magnitudes $\|\mathbf{x}\|$ and $\|\mathbf{p}\|$, so we can simplify the relation to

$$\mathbf{x} \quad = \quad L(\|\mathbf{p}\|)\mathbf{p}$$

By choosing a parametric approximation to the true function L for a given camera, we may convert between image and pinhole coordinates. Calling the conversion from pinhole to image the *forward* transform, we will write the *inverse* transform in the form

$$\mathbf{p} \quad = \quad L^i(\|\mathbf{x}\|)\mathbf{x}$$

3. The division model for distortion

In order to render the mathematics tractable later in the paper, we will need a distortion approximation that differs from the normal model. Let us now derive the new model and show that it performs as well as the traditional approximation.

True lens distortion curves are typically very complex, and systems which deal carefully with nonlinear distortion [2, 19] use high-order models or lookup tables to calibrate their cameras. For computer vision, however, and particularly for *matching*, accuracies of the order of a pixel are all that are required. Thus, it is common to expand the distortion function L as a Taylor series, and to keep only the first nonlinear even term:

$$\mathbf{x} = (1 + \lambda\|\mathbf{p}\|^2)\mathbf{p} \tag{1}$$

However, other models have been proposed, and one in particular which I shall call the *division model* will prove useful later. Note that this alternative model is not an approximation to the more usual model, but a different approximation to the true curve. Figure 3 shows how they compare. The *division model* is written

$$\mathbf{p} = \frac{1}{1 + \lambda\|\mathbf{x}\|^2}\mathbf{x} \tag{2}$$

Now, it is crucial to remember that equation 2 is *not* an approximation to equation 1. Both are approximations to the camera's true distortion function. However, it is interesting to see how the new approximation compares to the old, and how both compare to calibrated ground truth. To validate the models against calibrated ground truth, the true distortion curve for a laboratory camera was obtained using a dense calibration grid placed directly in front of the lens. This curve is shown in Figure 3, along with the best fits of both approximations. The accuracy of both approximations is essentially the same: $\text{RMS}_{\text{old}} = 0.77$ pixels, $\text{RMS}_{\text{new}} = 0.65$ pixels. Thus, although this is just one experiment, we can be reasonably confident that the newer model is as good an approximation as the traditional one.

4. Algorithm 1: Linear estimation of F and λ

Let us now derive this paper's main contribution. It is known that in order to compute the fundamental matrix from perfect point correspondences $\mathbf{p} \leftrightarrow \mathbf{p}'$, we may use the 8-point algorithm [6]. In this section, we show how the 8-point algorithm can be modified to include λ, the distortion parameter, and thus compute F from distorted, measurable, points $\mathbf{x} \leftrightarrow \mathbf{x}'$.

4.1. Review: The 8-point algorithm

A point correspondence in pinhole coordinates $\mathbf{p} \leftrightarrow \mathbf{p}'$ which corresponds to a real 3D point which has been imaged by a pair of cameras will satisfy the epipolar constraint. This is embodied in the *fundamental matrix*, F, for the pair

$$\mathbf{p}'^{\top}\mathsf{F}\mathbf{p} = 0 \tag{3}$$

It is the task of this paper to recover F from point correspondences. Writing $\mathbf{p} = (p, q, 1)$, and concatenating the rows of F into a 9-vector $\mathbf{f}$, we may rewrite the above constraint as

$$[p'p, p'q, p', q'p, q'q, q', p, q, 1] \cdot \mathbf{f} = 0$$

Collecting 8 such rows into a design matrix, D, we obtain an estimate for $\mathbf{f}$ by solving $D\mathbf{f} = 0$. This estimate will be greatly improved [8] by truncating the resulting matrix to rank 2.

4.2. Incorporating the distortion parameter

In order to compute F from the known image coordinates $\mathbf{x}$, we must express (3) in terms of $\mathbf{x}$. Writing the new distortion equation (2) projectively, we obtain:

$$\begin{aligned}
\begin{pmatrix} p \\ q \\ 1 \end{pmatrix} &= \begin{pmatrix} x \\ y \\ 1 + \lambda(x^2 + y^2) \end{pmatrix} \\
&= \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} + \lambda \begin{pmatrix} 0 \\ 0 \\ \|\mathbf{x}\|^2 \end{pmatrix} \\
\therefore \mathbf{p} &= \mathbf{x} + \lambda\mathbf{z}
\end{aligned}$$

where both $\mathbf{x}$ and $\mathbf{z}$ are known (can be computed from image coordinates alone). Then the epipolar constraint is

$$\begin{aligned}
(\mathbf{x}' + \lambda\mathbf{z}')^{\top}\mathsf{F}(\mathbf{x} + \lambda\mathbf{z}) &= 0 \\
\mathbf{x}'^{\top}\mathsf{F}\mathbf{x} + \lambda(\mathbf{z}'^{\top}\mathsf{F}\mathbf{x} + \mathbf{x}'^{\top}\mathsf{F}\mathbf{z}) + \lambda^2\mathbf{z}'^{\top}\mathsf{F}\mathbf{z} &= 0
\end{aligned}$$

which is quadratic in λ and linear in F. Indeed, expanding everything out, we obtain (with $r = \|\mathbf{x}\|$, and $r' = \|\mathbf{x}'\|$)

$$\begin{aligned}
&[\, x'x \;\; x'y \;\; x' \;\; y'x \;\; y'y \;\; y' \;\; x \;\; y \;\; 1 \,] \cdot \mathbf{f} + \\
+\ \lambda\ &[\, 0 \;\; 0 \;\; x'r^2 \;\; 0 \;\; 0 \;\; y'r^2 \;\; xr'^2 \;\; yr'^2 \;\; r^2 + r'^2 \,] \cdot \mathbf{f} + \\
+\ \lambda^2\ &[\, 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; r'^2 r^2 \,] \cdot \mathbf{f} = 0
\end{aligned}$$

Gathering the three row vectors into three design matrices, we obtain the following quadratic eigenvalue problem (QEP) [17]:

$$(D_1 + \lambda D_2 + \lambda^2 D_3)\mathbf{f} = 0 \tag{4}$$

Such problems are analogous to standard 2nd order ODEs (replace λ with partial derivative operators), and efficient numerical algorithms are readily available, for example MATLAB provides the function `polyeig`. Appendix A shows how there are at most 10 solutions and in practice no more than 6 which are real.

5. Algorithm 2: Planar homography estimation

The preceding analysis applies also to the estimation of a plane projective transformation between the images. In this case, each point correspondence adds two rows (see [8, p71]) to the design matrices, viz.

$$\begin{aligned}
D_1 &= \begin{bmatrix} 0 & 0 & 0 & -x' & -y' & -1 & yx' & yy' & y \\ x' & y' & 1 & 0 & 0 & 0 & -xx' & -xy' & -x \end{bmatrix} \\
D_2 &= \begin{bmatrix} 0 & 0 & 0 & -rx' & -ry' & -r'-r & 0 & 0 & yr' \\ rx' & ry' & r'+r & 0 & 0 & 0 & 0 & 0 & -xr' \end{bmatrix} \\
D_3 &= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & -rr' & 0 & 0 & 0 \\ 0 & 0 & rr' & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

The analogous computation for the trifocal tensor leads to a cubic eigenvalue problem, which is again readily solved.

6. Implementation

The preceding section has presented a theoretical solution to the computation of the fundamental matrix and lens distortion. Immediately, the question of stability arises. Are the equations stable enough to be used in real-world problems? I shall show that the answer is yes, providing that a robust harness such as RANSAC is used for computation, and that care is taken with the use of the model to determine whether candidate points are inliers or outliers. In this sense, the kernel is very similar in performance to the successful 7-point algorithm for fundamental matrix computation [8].

6.1. Synthetic tests

In order to gain a feeling for the performance of the basic algorithm under typical image noise conditions, an investigation with synthetic data was conducted. A realistic scene was generated using 3D points and camera positions from standard point-based reconstruction on a low-distortion sequence (20 frames of Figure 1). These points and cameras were used to generate perfect 2D points, to which Gaussian noise was added. Because the 3D structure came from a real-world reconstruction, we may be confident that the arrangement is generic, while maintaining the control over system parameters that a synthetic test requires. The testing procedure was as follows.

1. Given 3D points $\{\mathbf{X}_i\}_{i=1}^N$ and 3×4 camera matrices P and P', generate two-view point correspondences $\{\mathbf{p} \leftrightarrow \mathbf{p}'_i\}_{i=1}^N$. Distort the perfect correspondences to generate noiseless image points $\tilde{\mathbf{x}}_i$. For this experiment, $N = 243$.

2. Repeat 100 times

   (a) Draw noise from a Gaussian distribution of standard deviation σ, and add to $\tilde{\mathbf{x}}_i$ giving noisy correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}_i{}'$.

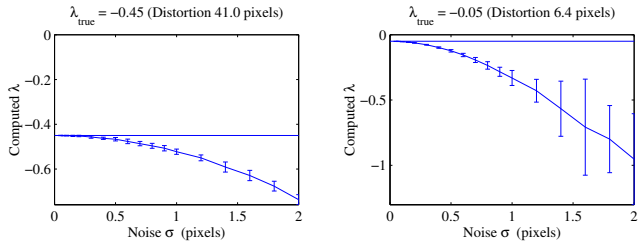


Figure 4: Tests on synthetic data. Synthetic image size is 640×480, and plausible 3D data are obtained by using stereo reconstruction from a real scene. The graphs show the computed distortion coefficient λ as a function of noise level on the 2D points. The horizontal line is the nominal value. A systematic bias to high corrections is evident, and is more pronounced on images with less distortion. Compare this to linear ellipse fitting.

(b) Scale image points by subtracting the image center and dividing by image diameter $(W + H)$.

(c) Form the $N \times 9$ design matrices D_1, D_2, D_3.

(d) Use MATLAB to compute $[V, \Lambda^{-1}] =$ `polyeig`$(D_1^\top D_3, D_1^\top D_2, D_1^\top D_1)$. Now V is the matrix of eigenvectors, and Λ^{-1} the vector of corresponding (inverse) eigenvalues.

(e) Discard imaginary, null, and infinite eigenvalues from Λ^{-1}, leaving 4–6 solutions. In this test, values where $|\lambda| > 10$ were also discarded.

(f) Store all remaining values of λ.

3. From the list of 100–600 computed λ values, compute the median and 10th and 90th percentile points. These are the values and errorbars in Figure 4.

The noise levels used had σ between 0 and 2 pixels. This represents a typical range in video and film imagery, with most cameras in our laboratory yielding σ of about 0.2 pixels after subpixel interest-point extraction.

Examining Figure 4 allows a number of conclusions about the algorithm to be drawn. Firstly, there is a systematic bias in the estimate of λ as noise level increases. The bias is towards more extreme distortions than the veridical value, and increases when distortion is small. For distortion of 40 pixels at the image corner, the estimate is within 20% of the veridical value at typical noise levels; for mild distortion (6 pixels), the estimate is many times the veridical value. Thus, this technique cannot give reliable estimates of lens distortion parameters when the amount of distortion is small.

This is similar to the pattern observed in linear solutions to the least-squares fitting of ellipses [9]. There, as the curvature of an elliptical arc decreases, and the arc approaches a line, the estimate of curvature is strongly biased. Of course, the curve still fits the data well, but the value of the curvature will be incorrect.

So, if the algorithm applies only when distortion is moderate to high, is it useful? Zhang's conclusion [20] is that one should use the pinhole camera for low-distortion images, and switch to a model including lens distortion if the pinhole model does not fit well. This is a reasonable strategy, and places the "pinhole+distortion" model on a continuum with the pinhole and affine camera approximations. However, the strategy does bring up the question of how to select an appropriate model. It might be preferable, if possible, to use the same model for all matching. This is only possible if the more general model does not overfit the data, accepting many bad matches rather than a smaller number of correct ones.

7. A robust algorithm for real data

In this section, we apply the model to the matching of real image pairs, and compare it to the pinhole model in cases of moderate and low distortion. In order to do so, however, we must incorporate the algorithm into a robust estimation strategy [8]. The input to such an algorithm is illustrated in Figure 5a. A pair of images is captured, interest points are computed, and each point in the first image is matched to each point within a window (of 100×100 pixels say) in the second image. These matches are limited to those pairs for which cross-correlation of image neighbourhoods is above a threshold. The set of matches produced by such a process is shown in Figure 5b.

Our input, then, is a set of noisy image correspondences $\mathbf{x} \leftrightarrow \mathbf{x}'$, some of which are incorrect. In the example shown, we have 513 correspondences, about 200 of which are correct. We shall use a RANSAC strategy to eliminate the incorrect correspondences and estimate the fundamental matrix and lens-distortion parameter. The most important component of such an algorithm is the test which is used to mark each correspondence as inlier or outlier. If the algorithm is to be effective, this test must measure a statistically meaningful quantity in image space in order that thresholds may be meaningfully set.

Without distortion, and assuming Gaussian[1] noise on the (x, y) positions of interest points, the optimal measure is given by

$$\epsilon(\mathbf{p}, \mathbf{p}', \mathsf{F}) = \min_{\{\hat{\mathbf{p}}, \hat{\mathbf{p}}' | \hat{\mathbf{p}}'^\top \mathsf{F} \hat{\mathbf{p}} = 0\}} \|\hat{\mathbf{p}}' - \mathbf{p}'\|^2 + \|\hat{\mathbf{p}} - \mathbf{p}\|^2 \quad (5)$$

A direct solution for this error in the pinhole case has been given by Hartley and Sturm [7]. In our case, this should be

[1] Although feature-point noise is not Gaussian, it does tend to be symmetric, so the above scoring function is monotonically related to the maximum likelihood.

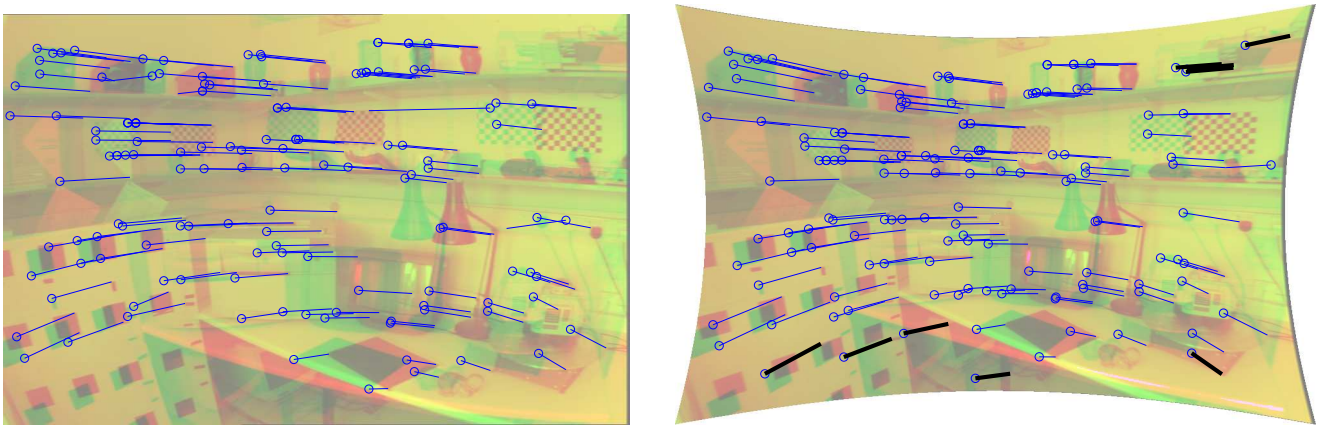


Figure 6: Example on real images. Matches selected as inliers for F (*Left*) and $\mathsf{F}\&\lambda$ (*Right*, distortion corrected via the calculated λ). F gives 130 matches plus 4 outliers, while the new model gives 140 matches, 1 outlier, and places the new matches (thicker tracks on the right) towards the periphery of the image.

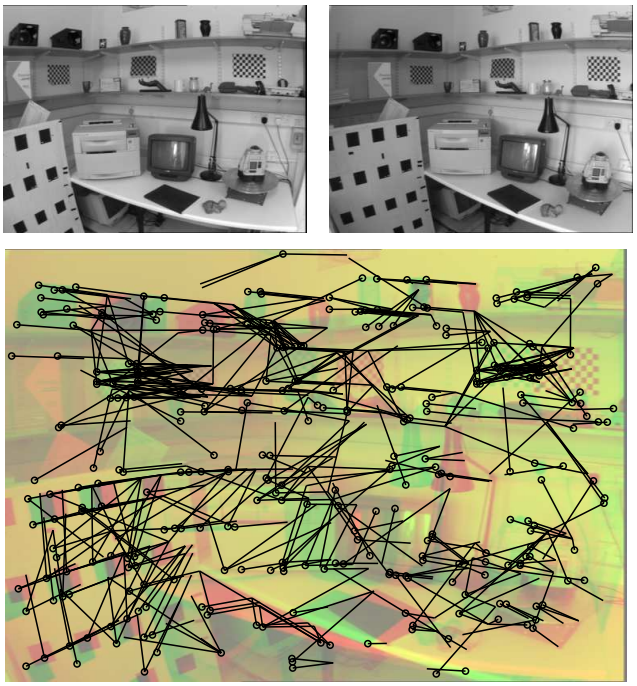


Figure 5: *Top (a)*: Input stereo pair. The lens is the same as that used in Figure 3. *Bottom (b)*: Matches input to the RANSAC algorithm. A circle marks a point in one image, lines join it to its potential matches in the next. The images are superimposed in different colour channels.

modified to also undistort the corrected points, thus:

$$\epsilon(\mathbf{x}, \mathbf{x}', \mathsf{F}, \lambda) = \min_{\{\hat{\mathbf{p}}, \hat{\mathbf{p}}' | \hat{\mathbf{p}}'^{\top} \mathsf{F} \hat{\mathbf{p}} = 0\}} \|\mathcal{L}(\hat{\mathbf{p}}') - \mathbf{x}'\|^2 + \|\mathcal{L}(\hat{\mathbf{p}}) - \mathbf{x}\|^2$$

However, we have chosen $\mathcal{L}$ to be easy to invert, but hard to compute, so this calculation looks impractical. The approximations which have been used in the pinhole case, such as distances from points to epipolar lines become the distance from points to curves [20] in our case.[2] On the other hand, we may not simply undistort the scene points and use ϵ, as this would be made small simply by setting λ to large positive values.

This paper uses a good approximation, which is to undistort the noisy image coordinates ($\mathbf{p} = \mathcal{L}^i(\mathbf{x})$), perform the Hartley-Sturm correction (i.e. find the minimizing $\hat{\mathbf{p}}'$ and $\hat{\mathbf{p}}$ from (5)), and then distort the corrected coordinates ($\hat{\mathbf{x}} = \mathcal{L}(\hat{\mathbf{p}})$). The distance between the corrected and original coordinates gives $\epsilon = \|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|^2 + \|\tilde{\mathbf{x}}' - \hat{\mathbf{x}}'\|^2$.

8. Experiments

The behaviour which is of greatest interest in the context of this paper is performance when compared to the current 7-point algorithm for matching. In the presence of lens distortion, current best practice is to run the 7-point algorithm with an artificially high acceptance threshold.

Figure 6 shows the subset of correspondences selected by the old and new kernels after 1000 RANSAC samples, on the image pair in Figure 5. The new algorithm finds more correspondences, and covers more of the scene, while the linear algorithm is limited by its inability to model distortion. Figure 7 shows that while increasing the old algorithm's acceptance threshold allows it to find more inliers, it also includes more false positives.

A combined test of robustness and tolerance of *small* amounts of distortion is provided by sequence "flowers" in

[2] Although, under the new distortion model, the epipolar curves are conics (indeed circles), for which closest point computation is easier than the cubic curves produced by the old model.

figure 8 where correspondences are correctly obtained on a natural scene with small distortion. On 40 image pairs, the average number of tracks was increased fractionally (1%), with an average computed λ of -0.10 and standard deviation $\sigma_\lambda = 0.09$. The remainder of Figure 8 shows three more 20-frame sequences on which performance is similar: "office" $\kappa = -0.06 \pm .11$, "gatehouse" $\kappa = -0.03 \pm 0.11$, and "production" $\kappa = -0.12 \pm 0.21$. The moderate-distortion sequences saw improvements in the number of 20-frame tracks of 4–5%, while the other low-distortion sequence ("office") saw an improvement from 190 to 193 tracks or 1.6%.

These results have the important consequence that we do not need to revert to a pinhole model for low-distortion scenes, allowing the proposed algorithm to be used by default rather than as a special case, and therefore allowing the construction of more general systems.

8.1. Known distortion center: implications

An assumption throughout this paper has been that the distortion center is known. In the absence of any other information, one would place it at the center of the image. Note that this does not place any constraint on where the *principal point* of the pinhole camera must be, and the two will in general be different [19]. Fixing the distortion center is a reasonable approximation for two reasons:

1. Although the principal point is an important camera calibration parameter, the precise positioning of the distortion center does not strongly affect the correction [19]. Indeed, including the distortion center in the calculations for Figure 3 improves the RMS by less than 0.03 pixels. Also, Figure 2 shows that simply including one term of image-centered correction is enough to greatly improve the performance of one real vision application.

2. Because we are estimating the *uncalibrated* geometry between a camera pair, our choice of coordinate system in the image plane has no effect on the resulting geometry (i.e. the fundamental matrix). Therefore fixing the distortion center has no negative implications on possible solutions for the *geometric* image center.

9. Summary and Conclusions

This paper has extended the uncalibrated estimation of geometry from multiple images to include a correction for lens distortion. The main contribution is a linear algorithm for the simultaneous estimation of a single lens distortion coefficient and the fundamental matrix. All previous algorithms have been iterative.

The paper has demonstrated that the algorithm has behaviour similar to other linear algorithms [8]—with systematic bias and moderate to poor tolerance to high noise—and, like other algorithms, works well as a computational kernel for robust estimators of two-view geometry.

The most significant speed implication of the new algorithm is in the number of RANSAC samples needed to ensure a certain degree of accuracy p. Each algorithm requires $\log(1-p)/\log(1-\epsilon^n)$, where ϵ is the proportion of data believed to be inlying, and $n = 7$ for the old algorithm, $n = 9$ for the new. The increase in the number of samples is then $\log(1-\epsilon^7)/\log(1-\epsilon^9) \approx \epsilon^{-2}$, or a factor of 4 for a 50%

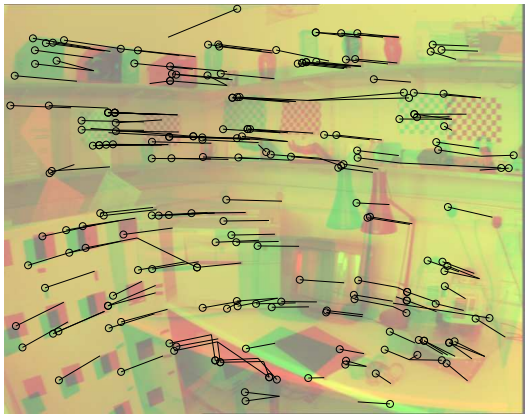
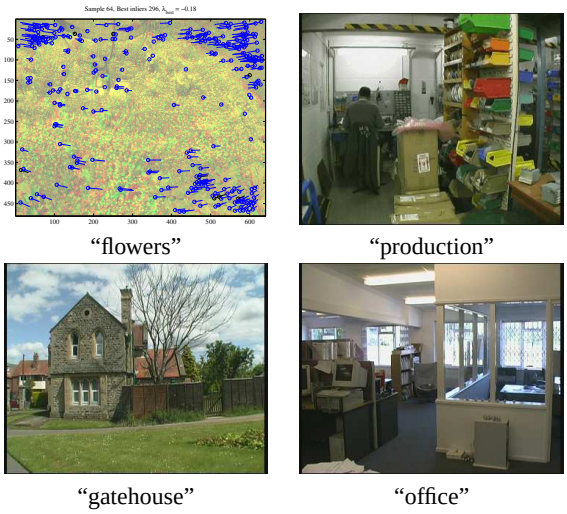


Figure 7: Using a artificially high threshold (4 pixels), the traditional fix for F estimation with distortion. More inliers are obtained but at the expense of including false matches.



"flowers" "production"

"gatehouse" "office"

Figure 8: Natural & indoor scenes, small to moderate distortion. (Top left:) The correct tracks indicate that the model is robust under conditions of high outlier percentage—the flowers were blowing in the wind—and low distortion. Tests were run on 41 frames (40 pairs) from this sequence. (Others:) Assorted sequences from which 60 further image pairs were taken for testing.

inlier percentage.

The important general conclusion is that it is now possible to match images which exhibit lens distortion with the same ease as those which accurately fit the pinhole model. Furthermore, one may use the distortion-aware model to match even low distortion images without overfitting.

A. Analysis of the QEP

This appendix shows how the number of solutions of (4) is reduced from 18, for the general case, to 10, and in practice six or fewer. For 9 points (the minimum for a solution), these matrices are square (9×9) and standard techniques may be applied. If more than 9 points are to be used to obtain a least-squares solution, premultiplication by $D_1^\top$ does not change the solution, but allows square solvers to be used.

To begin, define a new variable $\mathbf{u} := \lambda\mathbf{f}$, giving

$$D_1\mathbf{f} + \lambda D_2\mathbf{f} + \lambda D_3\mathbf{u} = 0 \tag{6}$$

Solving this system for $\mathbf{u}$ and $\mathbf{f}$ will obviously not solve the original problem, but solving *simultaneously* with $\mathbf{u} = \lambda\mathbf{f}$ will yield all the values of $\mathbf{f}$ which solve the original system. This *pair* of equations

$$\left.\begin{aligned} -D_1\mathbf{f} + 0\mathbf{u} &= \lambda D_2\mathbf{f} + \lambda D_3\mathbf{u} \\ 0_{9\times9}\mathbf{f} + I_{9\times9}\mathbf{u} &= \lambda I_{9\times9}\mathbf{f} + \lambda 0_{9\times9}\mathbf{u} \end{aligned}\right\}$$

may be written in block matrix form as follows:

$$\begin{bmatrix} -D_1 & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix}\mathbf{f}\\ \mathbf{u}\end{bmatrix} = \lambda\begin{bmatrix} D_2 & D_3 \\ I & 0 \end{bmatrix}\begin{bmatrix}\mathbf{f}\\ \mathbf{u}\end{bmatrix} \tag{7}$$

$$A\mathbf{v} = \lambda B\mathbf{v} \tag{8}$$

This 18×18 generalized eigensystem may be solved either by premultiplying by $\begin{pmatrix} -D_1^{-1} & 0 \\ 0 & I \end{pmatrix}$ to convert it to a regular unsymmetric eigensystem, or by using the QZ algorithm [5].

Because D_3 has eight all-zero columns, the matrix B has rank at most 10. Its nullvectors will correspond to infinite eigenvalues, as they will satisfy $\mu A\mathbf{v} = B\mathbf{v}$ for $\mu = 0$. All solutions $\mathbf{v}$ which correspond to non-infinite eigenvalues may be written $M\mathbf{z}$ for $M = \ker(\ker B)^\top$ and $\mathbf{z} \in \mathbb{R}^{10}$. Thus we may solve the 10×10 system $M^\top A M^\top \mathbf{z} = \lambda M^\top B M\mathbf{z}$. In practice, 4 of these 10 eigenvalues have been found to be imaginary, although a proof is future work. Degeneracies of the problem include (1) the set of degeneracies of F, which make D_1 singular, and (2) those which cause D_2 to drop rank. Again, characterization of these is further work.

Acknowledgments

Thanks to David Capel, Ben Tordoff, Fred Schaffalitzky and Andrew Zisserman for discussions relating to this work, and to the anonymous reviewers. Generous funding was provided by the Royal Society.

References

[1] 2d3 Ltd. Boujou, 2000. `http://www.2d3.com`.

[2] D.C. Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3):444–462, 1966.

[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, Jul 1995.

[4] F. Du and M. Brady. Self-calibration of the intrinsic parameters of cameras for active vision systems. In *Proc. CVPR*, 1993.

[5] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, MD, second edition, 1989.

[6] R. I. Hartley. In defence of the 8-point algorithm. In *Proc. ICCV*, pages 1064–1070, 1995.

[7] R. I. Hartley and P. Sturm. Triangulation. In *Proc. Conference Computer Analysis of Images and Patterns*, Prague, Czech Republic, 1995.

[8] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. CUP, Cambridge, 2000.

[9] K. Kanatani. Statistical bias of conic fitting and renormalization. *IEEE PAMI*, 16(3):320–326, 1994.

[10] S. B. Kang. Semiautomatic mathods for recovering radial distortion parameters from a single image. Technical Report CRL 97/3, Digital CRL, 1997.

[11] P. McLauchlan and A. Jaenicke. Image mosaicing using sequential bundle adjustment. In *Proc. BMVC.*, pages 616–625, 2000.

[12] H. S. Sawhney and R. Kumar. True multi-image alignment and its application to mosaicing and lens distortion correction. In *Proc. CVPR*, pages 450–456, 1997.

[13] C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA, USA, 4th edition, 1980.

[14] G. P. Stein. *Geometric and Photometric Constraints: Motion and Structure from three Views*. PhD thesis, MIT, 1997.

[15] G. P. Stein. Lens distortion calibration using point correspondences. In *Proc. CVPR*, 1997.

[16] R. Swaminathan and S.K. Nayar. Nonmetric calibration of wide-angle lenses. In *Proc. CVPR*, pages 413–419, 1999.

[17] F. Tisseur and K. Meerbergen. The quadratic eigenvalue problem. *SIAM Review*, Jun 2001.

[18] W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, LNCS. Springer Verlag, 2000.

[19] R. G. Willson and S. A. Shafer. What is the center of the image? In *Proc. CVPR*, pages 670–671, 1993.

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