

Minimal Solutions for Panoramic Stitching

Matthew Brown
Microsoft Research
brown@microsoft.com

Richard I. Hartley
Australian National University and NICTA*
richard.hartley@anu.edu.au

David Nistér
Microsoft Research
dnister@microsoft.com

Abstract

This paper presents minimal solutions for the geometric parameters of a camera rotating about its optical centre. In particular we present new 2 and 3 point solutions for the homography induced by a rotation with 1 and 2 unknown focal length parameters. Using tests on real data, we show that these algorithms outperform the standard 4 point linear homography solution in terms of accuracy of focal length estimation and image based projection errors.

1. Introduction

Image stitching is the process of combining data from multiple images to form a larger composite image or mosaic. This is possible when the amount of parallax between the images is small or zero. This occurs, for example, when viewing a planar scene, or when rotating about a point.

Recent successful approaches to image stitching have used feature based techniques [2, 3]. These methods typically employ random sampling algorithms such as RANSAC for robust estimation of the image geometry, to cope with noisy and outlier contaminated feature matches [4]. The RANSAC step has an inner loop consisting of a fast solution for the parameters of interest given a small number of correspondences. Since the probability of choosing a set of correct correspondences decreases rapidly as the sample size increases, solutions that use as few points as possible in the RANSAC loop are favourable.

Recently, solutions have been developed that enable several problems in image geometry to be solved efficiently using the theoretical minimum number of points [12, 15]. These algorithms use algebraic geometry techniques [14] to solve directly for the parameters of interest using polynomial equations, and have been demonstrated to be superior to previous approaches in structure and motion problems [12]. Previous (non-minimal) approaches involved linear solutions for over-parameterised matching relations e.g. the

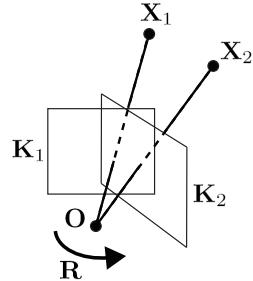


Figure 1: A pair of correspondences in a rotating camera gives a single constraint on the calibration matrices $\mathbf{K}_1, \mathbf{K}_2$. This paper describes solutions for the rotation and focal length of a pair of cameras given the minimal number of correspondences between them.

8 point DLT algorithm for the fundamental matrix [8], or the linear solution for the trifocal tensor [6]. In addition to being more noise prone, non-minimal solutions typically require a subsequent auto-calibration stage to estimate the minimal set of transformation parameters from the linear solution [18, 13]. This would be required for example, if one wanted to find optimal estimates for the underlying parameters using an iterative algorithm such as bundle adjustment [19].

The same arguments transfer to the analogous problems in image stitching. Though state of the art approaches currently use a 4 point linear solution for a homography in the RANSAC loop [8], we show that in the most common practical cases 2 or 3 points are in fact sufficient, and that the resulting solutions are more noise tolerant and give better initialisations for the underlying parameters.

We focus on cases where all the parameters other than the focal length are known. We believe that these are the most important practical cases as sensible priors often exist for the other parameters (e.g. central principal point, zero skew). Also, the resulting algorithms will use fewer points than the standard 4 point homography. We do not investigate the case of calibrated rotation where closed form solutions already exist [9].

*NICTA is a research centre funded by the Australian Government's Department of Communications, Information Technology and the Arts and the Australian Research Council.

A linear method for autocalibration of a rotating camera was first proposed by Hartley [5, 7]. We compare our minimal solutions to this technique in Section 10.

2. Problem Statement

Consider two cameras with coincident optical centres viewing a world point \mathbf{X}

$$\tilde{\mathbf{u}}_1 \sim \mathbf{K}_1 \mathbf{X} \quad (1)$$

$$\tilde{\mathbf{u}}_2 \sim \mathbf{K}_2 \mathbf{R} \mathbf{X} \quad (2)$$

where \mathbf{R} is the rotation between the cameras, $\mathbf{K}_1, \mathbf{K}_2$ are upper triangular calibration matrices and $\tilde{\mathbf{u}}_1, \tilde{\mathbf{u}}_2$ are the projections of the 3D point \mathbf{X} to images 1 and 2. By eliminating \mathbf{X} we find two constraints on the unknown parameters of \mathbf{R} (3 parameters) and $\mathbf{K}_1, \mathbf{K}_2$ (up to 5 parameters each). The goal is to solve for the rotation and intrinsic parameters of the cameras using the minimum number of correspondences possible.

3. Linear Algorithms

From equations (1) and (2) we can write

$$\tilde{\mathbf{u}}_2 \sim \mathbf{H}_{12} \tilde{\mathbf{u}}_1 \quad (3)$$

where

$$\mathbf{H}_{12} \sim \mathbf{K}_2 \mathbf{R} \mathbf{K}_1^{-1} \quad (4)$$

Eliminating \mathbf{R} gives

$$\mathbf{H}_{12} \omega_1^* \mathbf{H}_{12}^T \sim \omega_2^* \quad (5)$$

where

$$\omega_i^* \sim \mathbf{K}_i \mathbf{K}_i^T \quad (6)$$

ω_i^* is known as the dual image of the absolute conic (DIAC) [8]. Linear algorithms for autocalibration proceed by computing a general 3×3 matrix \mathbf{H}_{12} from 4 point correspondences [7]. The symmetric system of equations (5) then gives 5 linear constraints on the elements of ω_1^* and ω_2^* . \mathbf{K}_1 and \mathbf{K}_2 can be found by Cholesky decomposition of ω_1^* and ω_2^* .

4. Polynomial Algorithms

Instead of linearising we can work directly with the sought after parameters of \mathbf{R} and \mathbf{K} . Consider a pair of correspondences $\mathbf{u}_{11} \leftrightarrow \mathbf{u}_{21}, \mathbf{u}_{12} \leftrightarrow \mathbf{u}_{22}$ where

$$\tilde{\mathbf{u}}_{ij} \sim \mathbf{K}_i \mathbf{R}_i \mathbf{X}_j \quad (7)$$

We can eliminate \mathbf{R}_i by considering the angle between \mathbf{X}_1 and \mathbf{X}_2

$$\cos^2 \theta = \frac{(\mathbf{X}_1^T \mathbf{X}_2)^2}{|\mathbf{X}_1|^2 |\mathbf{X}_2|^2} \quad (8)$$

Noting that this angle is preserved in both images

$$\cos^2 \theta_1 = \cos^2 \theta_2 \quad (9)$$

and writing in terms of the camera parameters gives

$$\frac{(\tilde{\mathbf{u}}_{11}^T \omega_1 \tilde{\mathbf{u}}_{12})^2}{\tilde{\mathbf{u}}_{11}^T \omega_1 \tilde{\mathbf{u}}_{11} \tilde{\mathbf{u}}_{12}^T \omega_1 \tilde{\mathbf{u}}_{12}} = \frac{(\tilde{\mathbf{u}}_{21}^T \omega_2 \tilde{\mathbf{u}}_{22})^2}{\tilde{\mathbf{u}}_{21}^T \omega_2 \tilde{\mathbf{u}}_{21} \tilde{\mathbf{u}}_{22}^T \omega_2 \tilde{\mathbf{u}}_{22}} \quad (10)$$

where $\omega_i^{-1} \sim \omega_i^* \sim \mathbf{K}_i \mathbf{K}_i^T$. This gives a single non-linear constraint on the unknown parameters of $\mathbf{K}_1, \mathbf{K}_2$, and is used as a building block for our polynomial solvers.

5. The 2 Point Algorithm

A simple case arises when only the focal length is unknown, but common to the 2 views. This occurs frequently in practice, for example, when capturing a panorama without zooming. Substituting

$$\mathbf{K}_i = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (11)$$

in equation (10) gives

$$\frac{(a_{12} + f^2)^2}{(a_1 + f^2)(a_2 + f^2)} = \frac{(b_{12} + f^2)^2}{(b_1 + f^2)(b_2 + f^2)} \quad (12)$$

where

$$\begin{aligned} a_{12} &= \mathbf{u}_{11}^T \mathbf{u}_{12}, & a_1 &= |\mathbf{u}_{11}|^2, & a_2 &= |\mathbf{u}_{12}|^2 \\ b_{12} &= \mathbf{u}_{21}^T \mathbf{u}_{22}, & b_1 &= |\mathbf{u}_{21}|^2, & b_2 &= |\mathbf{u}_{22}|^2 \end{aligned} \quad (13)$$

Equation (12) can be simplified to give a 3rd degree polynomial in f^2 (the fourth degree term cancels). This can be solved in closed form.

6. The 3 Point Algorithm

Another important practical case occurs when the focal length is unknown and potentially different between the two views. This might occur when capturing a panorama with variable zoom. Substituting

$$\mathbf{K}_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (14)$$

in equation (10) gives

$$\frac{(a_{12} + f_1^2)^2}{(a_1 + f_1^2)(a_2 + f_1^2)} = \frac{(b_{12} + f_2^2)^2}{(b_1 + f_2^2)(b_2 + f_2^2)} \quad (15)$$

This can be simplified to give a biquadratic equation in f_1^2 and f_2^2 . Writing this as a quadratic in f_1^2 gives

$$pf_1^4 + qf_1^2 + r = 0 \quad (16)$$

where

$$\begin{aligned} p &= f_2^2(b_1 + b_2 - 2b_{12}) + b_1b_2 - b_{12}^2 \\ q &= f_2^4(2a_{12} - a_1 - a_2) + f_2^2(2a_{12}(b_1 + b_2) \\ &\quad - 2b_{12}(a_1 + a_2)) + 2a_{12}b_1b_2 - (a_1 + a_2)b_{12}^2 \\ r &= f_2^4(a_{12}^2 - a_1a_2) + f_2^2(a_{12}^2(b_1 + b_2) - 2a_1a_2b_{12}) \\ &\quad + a_{12}^2b_1b_2 - a_1a_2b_{12}^2 \end{aligned} \quad (17)$$

Note that p is 1st degree and q and r are second degree polynomials in f_2^2 . Taking 2 pairs of points gives the simultaneous equations

$$p_1f_1^4 + q_1f_1^2 + r_1 = 0 \quad (18)$$

$$p_2f_1^4 + q_2f_1^2 + r_2 = 0 \quad (19)$$

The common solutions of the polynomials are the roots of the resultant R

$$R = \begin{vmatrix} p_1 & q_1 & r_1 & 0 \\ 0 & p_1 & q_1 & r_1 \\ p_2 & q_2 & r_2 & 0 \\ 0 & p_2 & q_2 & r_2 \end{vmatrix} \quad (20)$$

This is a 7th degree polynomial in f_2^2 , which can be solved by taking the eigenvalues of the companion matrix. f_1^2 is then obtained by solving the quadratic equation (16).

Note that this is not strictly a minimal solution. The equations (18) and (19) are formed by choosing two pairs of correspondences from 3 point correspondences in 2 views. There are 3 ways to do this, which overconstrains the solution. Indeed 3 point correspondences gives 6 equations in only 5 unknowns, so the problem is overdetermined.

A neater solution with a lower degree polynomial can be obtained by using all 3 pairwise constraints from the 3 point correspondences

$$\begin{bmatrix} p_1 & q_1 & r_1 \\ p_2 & q_2 & r_2 \\ p_3 & q_3 & r_3 \end{bmatrix} \begin{bmatrix} f_1^4 \\ f_1^2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (21)$$

Since the determinant of this 3×3 matrix must vanish, we can write down a 5th degree polynomial in f_2^2 . Note that this solution enforces the fact that the 3×3 matrix has a nullvector, but discards the constraint that the nullvector is of the form $[x^2 \ x \ 1]^T$.

7. Other Algorithms

The technique described in the previous two sections can be extended for arbitrary $\mathbf{K}_1, \mathbf{K}_2$. Given correspondences in 2 views, up to 5 intrinsic parameters can be solved for by writing down equation (10) for as many point pairs as needed. The resulting simultaneous polynomials can be solved using techniques from algebraic geometry [14].

8. Solving for Rotation

Given the calibration matrices $\mathbf{K}_1, \mathbf{K}_2$ it is straightforward to solve for the rotation from corresponding image rays. We use the method in [1] which involves taking the SVD of the correlation matrix \mathbf{C} of the corresponding rays

$$\mathbf{C} = \sum_i \hat{\mathbf{r}}_{2i}\hat{\mathbf{r}}_{1i}^T \quad (22)$$

where

$$\mathbf{r}_{ij} = \mathbf{K}_i^{-1}\tilde{\mathbf{u}}_{ij} \quad (23)$$

and $\hat{\mathbf{r}}_{ij} = \mathbf{r}_{ij}/|\mathbf{r}_{ij}|$. Given that the SVD of the correlation matrix $\mathbf{C} = \mathbf{U}\Sigma\mathbf{V}^T$ the (least squares) estimate of the rotation is given by

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{bmatrix} \mathbf{V}^T \quad (24)$$

where

$$s = \text{sgn}(\det(\mathbf{UV}^T)) \quad (25)$$

Note that this is not a minimal solution as 2 point correspondences are used to solve for 3 rotation angles, but the solution minimises the squared distance between rays

$$\mathbf{R} = \arg \min_{\mathbf{R}'} \sum_i |\mathbf{r}_{2i} - \mathbf{R}'\mathbf{r}_{1i}|^2 \quad (26)$$

9. Evaluation of Solvers

In this section we describe our methodology for evaluation and comparison of the linear and polynomial solvers. For each solver, we use MLESAC [17] with n trials to estimate the pairwise motion parameters. For the linear solvers we use equation (5). Note that it is important to use a robust solution to these equations (see Appendix A). For the minimal solvers we use the algorithms of sections 5 and 6.

We use two evaluation functions, the rms focal length error and rms projection error. Both are evaluated over all image pairs, and against a gold standard solution obtained by bundle adjustment over the whole sequence as in [2]. Note that it is difficult to solve for the focal length even using bundle adjustment unless the panorama wraps around

360° [10], so we ensure that this is the case for both of our datasets.

The rms pixel error e_p is based on the projection errors of random image points, relative to the gold standard. Conceptually, we randomly distribute N points on each image i and use the gold standard and test homographies \mathbf{H}_{ij}^* , \mathbf{H}_{ij} to project the points to image j . The residual is the difference between the projections, and the error function a robust rms value of the residuals. See figure 2.

Note that we deliberately choose uniformly distributed points for this test, and not those found using some feature detector. This eliminates the choice of feature detector from the evaluation scores, and forces the solutions to be accurate even in areas where the feature detectors would not fire e.g. blank areas of sky.

More formally

$$e_p^2 = \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^n \sum_{\substack{k=1 \\ j \neq i}}^n \rho(\mathbf{r}_{ij}^{(k)}) \quad (27)$$

where $\mathbf{r}_{ij}^{(k)}$ is the difference between the projections of the k th random point from image i to image j under the gold standard and test homographies

$$\mathbf{r}_{ij}^{(k)} = \mathbf{u}_j^{(k)} - \mathbf{u}_j^{*(k)} \quad (28)$$

$$\tilde{\mathbf{u}}_j^{(k)} = \mathbf{H}_{ij} \tilde{\mathbf{u}}_i^{(k)} \quad (29)$$

$$\tilde{\mathbf{u}}_j^{*(k)} = \mathbf{H}_{ij}^* \tilde{\mathbf{u}}_i^{(k)} \quad (30)$$

$\mathbf{u}_i^{(k)}$ is a random point uniformly distributed in image i , and $\mathbf{H}_{ij}^* = \mathbf{K}_j^* \mathbf{R}_{ij}^* \mathbf{K}_i^{*-1}$ and $\mathbf{H}_{ij} = \mathbf{K}_j \mathbf{R}_{ij} \mathbf{K}_i^{-1}$ are the gold standard and test homographies from image i to image j . The third summation in equation (27) is over all randomly generated points $\mathbf{u}_i^{(k)}$ that successfully project inside image j under either the gold standard or test homographies, the number of such points is N . We use a robust error function

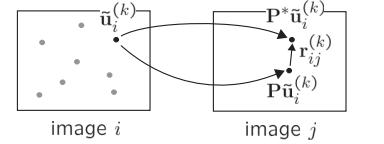
$$\rho(\mathbf{x}) = \begin{cases} |\mathbf{x}|^2, & \text{if } |\mathbf{x}| < \epsilon \\ \epsilon^2, & \text{if } |\mathbf{x}| \geq \epsilon \end{cases} \quad (31)$$

with a value $\epsilon = 10$ pixels used in all of our experiments.

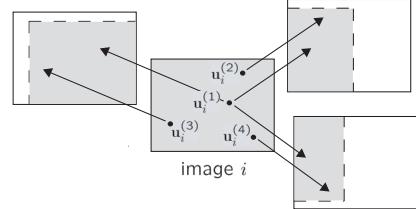
The rms focal length error e_f is given by

$$e_f^2 = \frac{\sum_i \sum_j \rho(f_{ij} - f_i^*)}{\sum_i \sum_j 1} \quad (32)$$

where f_{ij} is the estimate of the focal length of image i from the pairwise match between images i and j , and f_i^* is the gold standard focal length for image i obtained from bundle adjustment. Again we use a robust error function to suppress gross errors in f , with a value $\epsilon = \bar{f}^* = 800$ pixels.



(a)



(b)

Figure 2: Evaluation function based on rms pixel error. Uniformly distributed points $\mathbf{u}_i^{(k)}$ are projected from image i to each matching image j . The residual $\mathbf{r}_{ij}^{(k)}$ between the projections of the point under the gold standard and test homographies is computed (a). The error function is the root mean square of all such residuals. Random points are generated for all images i , and residuals are summed for all images that overlap image i (b).

10. Results

For evaluation we use 2 real image datasets – the first contains 42 images with common focal length. The second has 29 images where the focal length f varies by up to a factor of 2 and $\log(f)$ has a uniform distribution. See figures 7 and 8. We establish point matches between the images by finding matching SIFT features [11]. Each feature was matched to 6 nearest neighbours. This resulted in 108000 matches for the fixed focal length dataset, of which 69% were outliers. The variable focal length dataset had 309306 matches of which 93% were outliers. There were more outliers in the second dataset due to the differing scales of the images.

We ran our solvers on the datasets with variable number of MLESAC iterations and plot the results in figures 3 and 4. The error bars correspond to 2 standard deviations of the results over 10 runs. We found that the minimal solvers significantly outperformed their linear counterparts in terms of focal length estimation and the corresponding pixel errors. Note that the errors in these plots are caused by a combination of gross errors (bad MLESAC samples, failure of auto-calibration) as well as noise in the interest point positions.

When attempting to estimate f it is important to maximise a realistic likelihood function in the sampling process

(i.e. to use MLESAC instead of RANSAC). This is because the number of inliers can remain approximately constant across a large range of focal lengths. See figure 9. Our MLESAC likelihood function assumes that correspondences are i.i.d. with position errors distributed according to a heavy tailed Gaussian distribution $p(\mathbf{r}) \propto e^{-\rho(\mathbf{r})/\sigma^2}$, where $\rho(\mathbf{r})$ is a truncated quadratic as in equation (31). Note that summing log likelihoods gives a clear minimum at the correct focal length, whereas simply counting the number of inliers does not (see figure 9).

Figures 3 and 4 show that it is better to use minimal solutions if our goal is to estimate the focal length. However, in some applications computing an accurate f is not necessary, and only a set of consistent correspondences is required. This is often easier to achieve in practice. For example, autocalibration from \mathbf{H} may fail due to noise in some cases, even though the inliers to \mathbf{H} are correct. Also, a similarity model can be quite effective at finding consistent correspondences, without being able to estimate f . To clarify this issue we tested the minimal solutions against the homography and similarity transform models, counting only the number of inliers and without attempting to estimate the focal length. Figure 5 shows the results. Interestingly, the 2 point similarity model works well for small numbers of RANSAC iterations, but is overtaken by the 4 point homography as the number of iterations increases. Again, the minimal solutions give the best results.

Another point to note is that the minimal solutions are no more expensive to compute than their linear counterparts. For example, the 3 point algorithm requires solution of a 5×5 eigenvalue problem, which is less expensive to compute than the 8×9 SVD used in the 4 point DLT algorithm.

10.1. Initialisation for Bundle Adjustment

Our polynomial solvers are intended to be used in the first stage of stitching applications where the subsequent stages involve bundle adjustment. From RANSAC we have estimates of $\mathbf{R}_{ij}, \mathbf{K}_i, \mathbf{K}_j$ for every pair i, j , but to initialise the bundle adjuster we require a single set of parameters $\mathbf{R}_i, \mathbf{K}_i$ for each camera. Note that some of the pairwise estimates are noisy. For example, if the images have low overlap the focal length estimate is based on few points at the image edge. If the images have too high overlap the focal length may also be noisy – since zero rotation gives zero information about focal length.

Hence it is useful to have some method to combine rotation and focal length estimates from multiple views. A simple (RANSAC free) approach is to vote over all feature matches, e.g. for each randomly sampled pair of points we record a vote for the field of view of that image (see figure 8). Another approach we have used is to use the median of all pairwise focal length estimates (see figure 7). In both of these examples we have chained together pairwise rota-

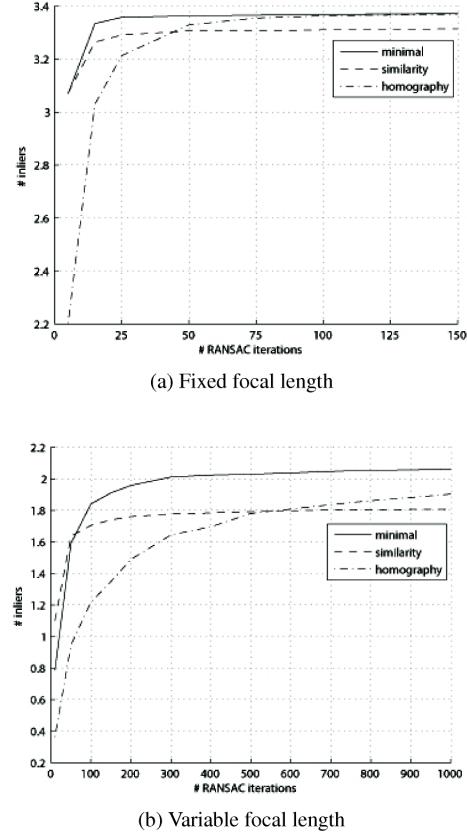


Figure 5: Comparison of minimal solutions, similarity transform and homography for finding inliers. Note that the minimal solutions are superior in both cases.

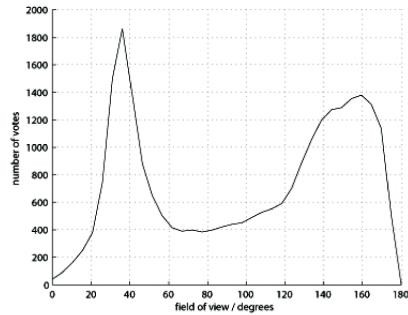


Figure 6: Votes for field of view from every random sample in the Eldorado sequence (fixed focal length). The correct field of view is around 40°. This value was used for the result of figure 8. This sequence contained 69% outliers.

tion estimates, each time adding the image with the most matches to the current set.

Despite our best efforts (minimal solutions and realistic

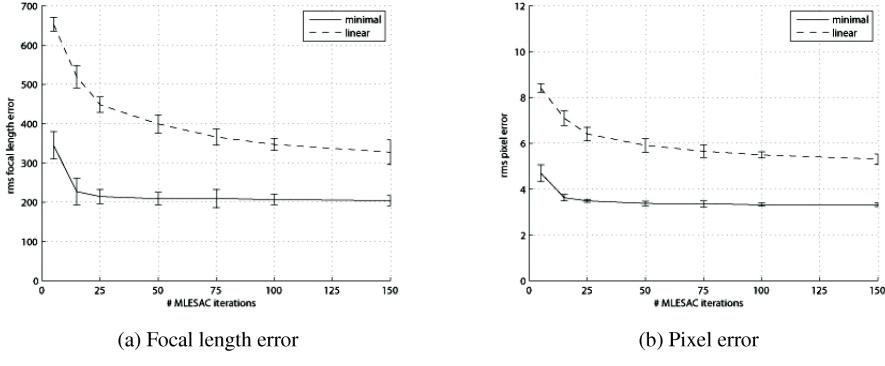


Figure 3: Comparison of linear and minimal solutions – fixed focal length.

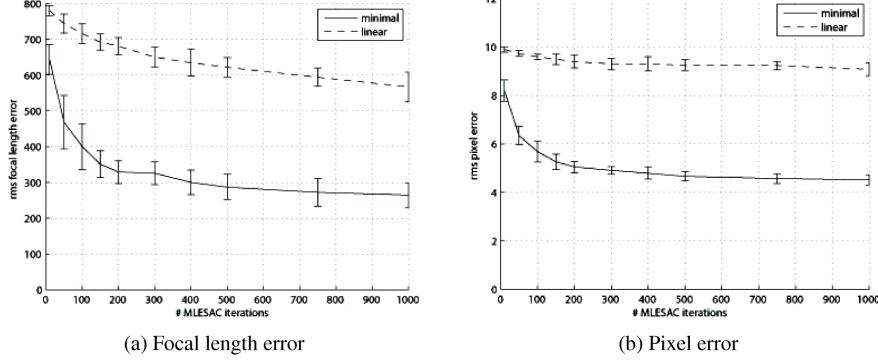


Figure 4: Comparison of linear and minimal solutions – variable focal length.

likelihood functions), there are limitations to what one can achieve using pairwise estimates of focal length. This is shown in figure 7(b), which shows the results of pairwise bundle adjustment on the variable focal length sequence. The overall scale of the focal lengths is still mis-estimated, despite using an optimal pairwise registration algorithm (the ends of the panorama should join up). Hence at some point, some form of global optimisation or bundle adjustment will be required (figure 7(c)).

11. Conclusions

We have presented a new family of minimal solutions for semi-calibrated panoramic stitching. This gives rise to simple 2 and 3 point algorithms for computing the focal length of a rotating camera. We have shown that these algorithms are more robust in terms of focal length estimation and rms pixel errors than previous state of the art solutions.

References

- [1] K. Arun, T. Huang, and D. Blostein. Least-squares fitting of two 3-D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:698–700, 1987.
- [2] M. Brown and D. Lowe. Recognising panoramas. In *Proceedings of the 9th International Conference on Computer Vision (ICCV03)*, volume 2, pages 1218–1225, Nice, October 2003.
- [3] M. Brown, R. Szeliski, and S. Winder. Multi-image matching using multi-scale oriented patches. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR05)*, San Diego, June 2005.
- [4] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography. *Communications of the ACM*, 24:381–395, 1981.
- [5] R. Hartley. Self-calibration from multiple views with a rotating camera. In *Proceedings of the 3rd European Conference on Computer Vision (ECCV94)*, pages 471–478, Stockholm, 1994.
- [6] R. Hartley. Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140, 1996.
- [7] R. Hartley. Self Calibration of Stationary Cameras. 22(1):5–23, 1997.



(a) Initialised with pairwise minimal solutions



(b) Initialised with pairwise bundle adjustment



(c) After full bundle adjustment and blending

Figure 7: Comparison of minimal solutions with pairwise bundle adjustment. Both initialisations converged to an rms error of less than 2 pixels after 6 iterations of global bundle adjustment. Note that the focal lengths for this sequence varied by up to a factor of 2.

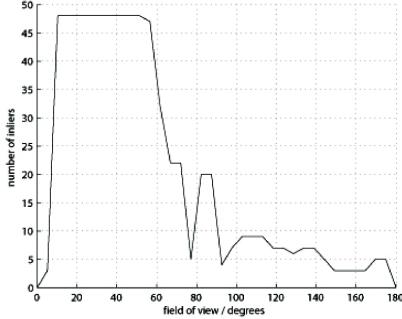


(a) Fixed focal length, without bundle adjustment, no blending

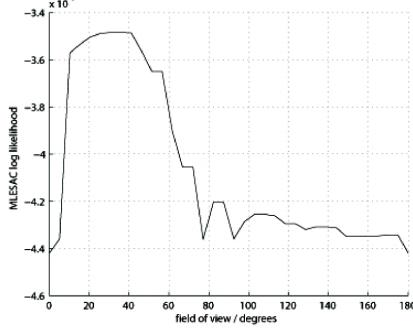


(b) Fixed focal length, with bundle adjustment, multi-band blending

Figure 8: Results for stitching without bundle adjustment. The focal length was initialised without RANSAC by using voting over all views (see figure 6).



(a) RANSAC



(b) MLESAC

Figure 9: These figures show the number of inliers (a) and MLESAC log likelihood (b) for a typical pair of images. Note that the MLESAC curve (log likelihood) shows a clear peak in focal length but the RANSAC curve (number of inliers) does not.

- [8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
- [9] B. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4:629, April 1987.
- [10] S. Kang and R. Weiss. Characterization of errors in compositing panoramic images. *Computer Vision and Image Understanding*, 73(2):269–280, February 1999.
- [11] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [12] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770, 2004.
- [13] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. *International Journal of Computer Vision*, 1999.
- [14] H. Stewénius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60:284–294, June 2006.
- [15] H. Stewénius, D. Nister, F. Kahl, and F. Schaffalitzky. A minimal solution for relative pose with unknown focal length. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR05)*, volume 2, pages 789–794, June 2005.
- [16] R. Szeliski and H. Shum. Creating full view panoramic image mosaics and environment maps. *Computer Graphics (SIGGRAPH'97)*, 31(Annual Conference Series):251–258, 1997.
- [17] P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, (78):138–156, 2000.
- [18] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR97)*, pages 609–614, 1997.

- [19] W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. In *Vision Algorithms: Theory and Practice*, number 1883 in LNCS, pages 298–373. Springer-Verlag, Corfu, Greece, September 1999.

A. Solving the Linear Equations

Given $\mathbf{H}_{ij} = s\mathbf{K}_j \mathbf{R} \mathbf{K}_i^{-1}$ equation 5 becomes

$$\mathbf{H}_{ij} \mathbf{K}_i \mathbf{K}_i^T \mathbf{H}_{ij}^T = s^2 \mathbf{K}_j \mathbf{K}_j^T \quad (33)$$

assuming $\mathbf{K}_i = \begin{bmatrix} f_i & 0 & 0 \\ 0 & f_i & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and expanding out the 6 equations of this symmetric system gives

$$f_i^2(h_{11}^2 + h_{12}^2) - s^2 f_j^2 + h_{13}^2 = 0 \quad (34)$$

$$f_i^2(h_{11}h_{21} + h_{12}h_{22}) + h_{13}h_{23} = 0 \quad (35)$$

$$f_i^2(h_{11}h_{31} + h_{12}h_{32}) + h_{13}h_{33} = 0 \quad (36)$$

$$f_i^2(h_{21}^2 + h_{22}^2) - s^2 f_j^2 + h_{23}^2 = 0 \quad (37)$$

$$f_i^2(h_{21}h_{31} + h_{22}h_{32}) + h_{23}h_{33} = 0 \quad (38)$$

$$f_i^2(h_{31}^2 + h_{32}^2) - s^2 + h_{33}^2 = 0 \quad (39)$$

In the case of common focal lengths $f_i = f_j$ one can find s by imposing the modulus constraint (that \mathbf{H}_{ij} has unit determinant). However, in this case we have found the equations involving s (34, 37 and 39) are noisy. Hence in both cases ($f_i = f_j$ and $f_i \neq f_k$) we have used a least squares solution of equations 35, 36 and 38 to solve for the focal length.