Minimal Solutions for Panoramic Stitching with Radial Distortion

Martin Byröd[1]
byrod@maths.lth.se

Matthew Brown[2]
mbrown@cs.ubc.ca

Kalle Åström[1]
kalle@maths.lth.se

[1] Centre for Mathematical Sciences
Lund University
Lund, Sweden

[2] Dept of Computer Science
University of British Columbia
Vancouver, Canada

Abstract

This paper presents a solution to panoramic image stitching of two images with co-inciding optical centers, but unknown focal length and radial distortion. The algorithm operates with a minimal set of corresponding points (three) which means that it is well suited for use in any RANSAC style algorithm for simultaneous estimation of geometry and outlier rejection. Compared to a previous method for this problem, we are able to guarantee that the right solution is found in all cases. The solution is obtained by solving a small system of polynomial equations. The proposed algorithm has been integrated in a complete multi image stitching system and we evaluate its performance on real images with lens distortion. We demonstrate both quantitative and qualitative improvements compared to state of the art methods.

1 Introduction

Given a sequence of images taken from a single point in space, but with varying orientations, it is possible to map the images into a common reference frame and create a perfectly aligned larger photograph with a wider field of view. This is normally referred to as panoramic image stitching. The main purpose of this paper is to extend previous work to account for camera distortion throughout the stitching process. This is in contrast to most previous approaches which have assumed a traditional pin-hole camera model. Stitching images with large radial distortion is useful in a practical context, as it allows the user to create 360 degree panoramas with wide angle lenses (often suffering from heavy radial distortion), using only a few exposures. Furthermore, radial distortion occurs frequently in both cheap consumer cameras and high-end lenses depending on the type of lens *e.t.c*.

In essence, a typical stitching pipeline consists of the following three parts

1. Image matching: Point matches across images are established and an initial estimate of the image geometry is computed. A RANSAC type algorithm is a popular choice here [10].

2. Bundle adjustment: The estimate of inner and outer calibration parameters is refined using non-linear optimization.
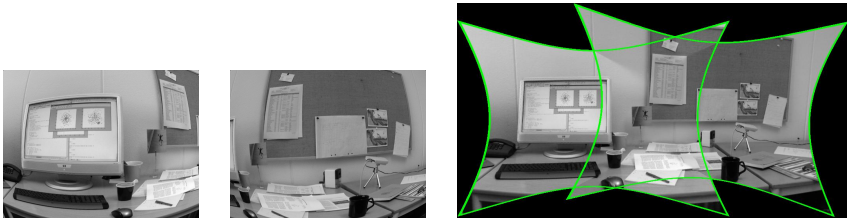
Figure 1: Left: Two images with heavy radial distortion taken with a common focal point. Right: The same two images after rectification and alignment using the stitching pipeline presented in this paper.

3. Rendering: The estimated camera parameters are used to project the images into a common reference frame.

This paper mainly deals with Step 1. At the core of the RANSAC loop is an algorithm which solves for calibration and geometry given a small set of corresponding points. Ideally one would like a solver which operates with the minimum possible number of correspondences. For instance, consider two images taken with a pin-hole camera calibrated up to focal length. We then need to estimate rotation (3 dof) and focal length (1 dof). Each point match yields two constraints, which means that the minimal solver should use two points. This problem was solved by Brown *et al.* in [3]. The rationale for using a minimal point set is that a smaller number of points yields a smaller probability of selecting a set contaminated by outliers. Furthermore, since we are solving directly for the parameters of interest, there is no need for an error-prone autocalibration process to extract the underlying camera parameters needed for multi-view non-linear techniques (i.e. bundle adjustment) to proceed.

The problem of dealing with cameras with various forms of non-linear distortions in computer vision is not new, but traditionally these lens effects have only been incorporated at the bundle adjustment stage. However, in situations with more than very mild distortions, it might be necessary to account for non-linear effects already at the RANSAC stage. In most cases, the dominant non-linear distortion effect is second order radial distortion. This effect was modeled in a neat way by Fitzgibbon [11] yielding an easy way to formulate and solve geometric computer vision problems with radial distortion. Following this, a number of contributions have been made on the estimation of geometric relations in computer vision along with radial distortion [1, 8, 15, 17]. This paper presents an efficient and stable solution to the problem of estimating rotation, focal length and radial distortion from two views with a common focal point. This is done using the minimal set of three point correspondences and the solution is obtained as one of the zeros of a system of polynomial equations. We demonstrate on real imagery that incorporating distortion already at the RANSAC stage yields a clear advantage.

1.1   Relation to Previous Work

The problem of image stitching is relatively well studied and a good overview of the literature and techniques can be found in the tutorial by Szeliski [21]. A complete stitching system representative of the state of the art in this area was presented by Brown and Lowe in [4].

A direct inspiration for this work is the two-point algorithm for estimating rotation and focal length by Brown *et al.* [3]. This algorithm does however not handle any distortion and we show that for non-standard lenses, this might be insufficient.

A related algorithm which *does* account for radial distortion due to Fitzgibbon [11] estimates homography and radial distortion using five correspondences. Two disadvantages of this approach is that (i) a homography is usually too general since in most cases one can assume square pixels and zero skew *e.t.c.* and (ii) the algorithm is not minimal. By contrast, our algorithm operates with three correspondences making it easier to find outlier-free sets.

Most closely related to our approach is the work by Jin [13]. Jin formulates the same problem as we do. However, Jin notes that solvers of polynomial equations are often numerically unstable and therefore abandons a direct solution approach. Instead he resorts to an iterative optimization based scheme. This is problematic since (i) convergence to a solution cannot be guaranteed (local minima) and (ii) even if a solution is found, this will only be one of the possible solutions and one cannot be sure to have found the right one. The actual problem has 18 (possibly complex) solutions and the only way to resolve this ambiguity is to test with additional points. Indeed, Jin reports poor performance of his algorithm for moderate to heavy distortions.

Fortunately, there has recently been progress in making polynomial solvers numerically robust [5, 6, 7, 8]. In this paper we make use of these techniques to provide a numerically stable true solver for the polynomial system, which is guaranteed to find all solutions.

2 Models for Panoramic Stitching

We consider a setup with two cameras P_1 and P_2 with a common focal point. We fix a coordinate system where the common focal point coincides with the origin and such that the first 3x3 part of the matrix P_1 is identity. Moreover, we have a set of world points $\{X_j\}$ and corresponding image projections $\{u_{1j}\}$ and $\{u_{2j}\}$. In most cases it is beneficial for stability to assume some partial calibration. A common choice is to assign square pixels, zero skew and centered principal point [12]. With this assumption we obtain the following relations

$$\lambda_{1j}u_{1j} = KX_j, \quad \lambda_{2j}u_{2j} = KRX_j, \tag{1}$$

where $K = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix}$, R is a rotation matrix and the λs are the depths. By normalizing to remove the dependence on λ_{ij} and solving for X_j we can write down the constraints

$$\frac{\langle K^{-1}u_{1j}, K^{-1}u_{1k}\rangle^2}{|K^{-1}u_{1j}|^2|K^{-1}u_{1k}|^2} = \frac{\langle X_j, X_k\rangle^2}{|X_j|^2|X_k|^2} = \frac{\langle K^{-1}u_{2j}, K^{-1}u_{2k}\rangle^2}{|K^{-1}u_{2j}|^2|K^{-1}u_{2k}|^2}, \tag{2}$$

where R has vanished from the right hand side since the scalar products and norms are invariant to rotations. The expression is squared to remove the square roots from the vector norms in the denominators. In the above equation f only occurs in even powers and hence we set $p = f^2$. Moreover we multiply through with p^2 to remove any $1/p^2$ terms. Finally we multiply up the denominators. At a first glance, this seems to yield a 4th degree polynomial in p but the 4th degree terms cancel out leaving a 3rd degree polynomial in p. This formulation was used in [3] to solve for the focal length. We next show how to modify this expression to include radial distortion.

2.1 A Three Point Minimal Solution for Distortion and Focal Length

Let x denote measured image coordinates affected by radial distortion and let u denote the corresponding pin-hole coordinates. We model radial distortion using Fitzgibbon's division

model

$$|x| = (1 + \lambda |x|^2)|u|, \tag{3}$$

where $|\cdot|$ is the vector length and λ is the radial distortion coefficient. This form has the advantage that in homogeneous coordinates we can write

$$u \sim x + \lambda z, \tag{4}$$

where $z = \begin{bmatrix} 0 & 0 & x_1^2 + x_2^2 \end{bmatrix}^T$.

We now simply insert (4) into (2) and obtain a polynomial of degree 3 in p and degree 6 in λ (the 8th and 7th degree terms in λ cancel out). One more unknown means that we need more constraints. With an additional point we can form three independent constraints of type (2). This situation is actually a little unsatisfactory since we cannot make use of all available information. Using all three constraints would yield an overdetermined system and hence there would be no solution in general. One possibility would be to introduce an extra unknown, but we found no natural way to do this and instead settled for selecting two of the three constraints to get a system of two equations in two unknowns. The experiments confirm that this strategy works well. We used the computer algebra software Macaulay2 [2] to check solvability and number of solutions for the system which is 18 in this case.

In addition to the problem formulation studied in this paper one could consider alternative setups with *e.g.* varying focal lengths and/or distortions. However, due to lack of space we will have to postpone a discussion of these to future publications and simply remark that we have studied alternative setups to some extent and so far found the case presented here to be the most practically useful.

3   Numerical Solution using Gröbner Basis Techniques

Solving a system of polynomial equations is often algorithmically difficult and there exist no efficient general purpose methods. Instead, specialized solvers are developed for particular cases. Recently, progress has been made using Gröbner basis techniques [5, 9, 19]. These solvers are known to suffer from numerical problems [16, 20] in some cases, but fortunately progress has been made on this point [5, 8]. Here, we make use of these advances to obtain an efficient and robust solver for the equations derived above. Due to lack of space we are not able to give a self contained treatment of the polynomial techniques and this section should be read in conjunction with *e.g.* [5, 19] for the method to be fully repeatable.

Gröbner bases are a concept within algebraic geometry, which is the general theory of multivariate polynomials over any field. See e.g. [9] for a good introduction to the field.

The goal is to find the set of solutions to a system $p_1(\mathbf{x}) = \cdots = p_m(\mathbf{x}) = 0$ of m polynomial equations in s variables. The polynomials $p_1, \ldots, p_m$ generate an *ideal I* in $\mathbb{C}[\mathbf{x}]$, the ring of multivariate polynomials in $\mathbf{x} = (x_1, \ldots, x_s)$ over the complex numbers. The Gröobner basis method for polynomial solving is based on computing the eigenvalues of multiplication mappings in the quotient space $\mathbb{C}[\mathbf{x}]/I$ in analogy with the companion matrix method for one variable polynomials. The key steps are to (i) expand the system of equations by multiplying with a problem dependent set of monomials and then write this on matrix form as a product between a coefficient matrix $\mathbf{C}$ and a vector of monomials $\mathbf{X}$, $\mathbf{CX} = 0$, and thereafter (ii) to put the system on reduced row echelon form using *e.g.* Gauss Jordan elimination which allows one to extract the necessary multiplication mappings.

In the current application, we are faced with a system of two equations in two unknowns (f, λ) occurring up to degrees 3 and 6 respectively and 18 solutions. We order the monomials

in *grevlex* order and multiply the two equations with all possible monomials up to degree 8, yielding a 90×132 coefficient matrix $\mathbf{C}$.

With the straightforward method of [19] we were not able to solve the problem and we had to employ a technique referred to as the *redundant solving basis method* in [5]. With this method one "pretends" to have system with more solutions which is easier to solve. This produces all the right solutions along with a set of false solutions which have to be filtered out by evaluation in the original equations. By using this technique and setting the solution set to 25 zeros (18 for the true system), we were able to get a stable solution.

An interesting comparison would be to run the automatic solver generator by Kukelova *et al.* [14]. This solver does not include any of the stabilizing methods mentioned above and might therefore fail, but this is yet to be investigated.

The algorithm has been implemented in MATLAB, which is not ideal for speed. However, the running time is dominated by an LU factorization and an eigenvalue decomposition which are fast in MATLAB so our implementation should not be too far behind a fully native implementation. The running time is about 13 milliseconds/instance on a standard 2 *Ghz* machine. The code is available for download at http://www.maths.lth.se/vision/downloads.

4 System Overview

The image stitching system implemented for this paper follows the typical pattern of modern geometric computer vision systems. We start off by finding matching points pair wise across images using the SIFT descriptor/detector [18] together with RANSAC for outlier rejection [10]. Thereafter we perform first a pair-wise and subsequently a global bundle adjustment step to get an accurate estimate of geometry and calibration parameters [12]. Finally we render the images onto an enclosing cylinder which can be cut and unfolded to the final panoramic image.

5 Experiments

In this section, we study the basic properties of the new algorithm on synthetic data and also assess its performance as part of a complete stitching system. For this purpose we have collected two data sets using a lens with significant non-linear distortion. The data sets referred to as *City* and *University* consist of 9 and 10 photographs respectively and both cover 360 degrees. In addition, a reference set called *Canal* consisting of 8 images was shot with a low distortion lens. The final result after matching, bundle adjustment and basic blending is shown in Figure 6. In all cases with image data we normalized the pixel coordinates to make the width of the image fall in the interval $[-1, 1]$. This makes values for λ independent of image resolution.

5.1 Robustness to Noise

We first did a basic sanity check of our new algorithm on synthetic data to study its behavior under noise compared to Fitzgibbons five-point algorithm for homography and distortion. Since Fitzgibbons algorithm estimates more degrees of freedom than needed to express a transformation with focal length, rotation and distortion, we expect to see some more sensitivity to noise than with our exact solver. For this experiment we randomly generated two views separated by a random rotation and drew three and five world points respectively
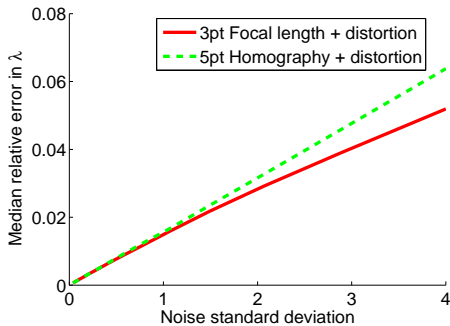
Figure 2: Error versus noise on synthetic data for the new three point algorithm and the five point algorithm for distortion and homography. Despite being over determined, the five point algorithm shows a slightly larger sensitivity to noise, probably due to the fact that the underlying model has more degrees of freedom.

from a normal distribution. The points were projected into the two views to form image point correspondences and a distortion of $\lambda = -0.5$ was applied. Finally varying degrees of noise (equivalent to the interval 0 to 4 pixels in an 800 pixels wide image) was added to the projected coordinates and the distortion parameter was estimated using each algorithm. This experiment was repeated 10000 times for each noise level and median errors were calculated. The median error was chosen since both algorithms (and in particular the five point algorithm) occasionally produce gross errors for unfortunate point configurations. This makes the average errors uninformative. The results are shown in Figure 5.1. As expected, both algorithms work well at low noise levels, but the five-point algorithm is slightly less robust at high noise levels.

5.2   Relation to Jin's Work

Since the work of Jin [13] is most closely related to the work presented here, a direct comparison would have been ideal, but we have not been able to obtain an implementation of Jin's method which is a little unsatisfactory. However, this should not be too serious, since under the assumption that Jin's method finds the right solution, the results should be virtually identical to ours. The problem is that Jin's method is *not* guaranteed to produce the desired solution. Figure 1 in [13] shows statistics of how often Jin's algorithm finds the correct solution. For distortions below -0.2 this rate is down to below 40%. In comparison, our solver is guaranteed to find the right solution for all distortions with no serious sacrifice in speed.

5.3   Performance in RANSAC

The main motivation for the three point algorithm presented in this paper is that it can be used to improve the RANSAC part in a stitching pipeline. With a refined inner step for geometry estimation, we hope to recover a larger proportion of inliers, at a higher rate and to a higher precision. In the next experiment we study the rate at which inliers are discovered as the RANSAC loop progresses. In addition to the five point algorithm, we now also compare our algorithm to the two-point algorithm of Brown *et al*. [3], which solves for focal length but not distortion. Brown *et al*. show in [3] that their algorithm is superior to the standard four
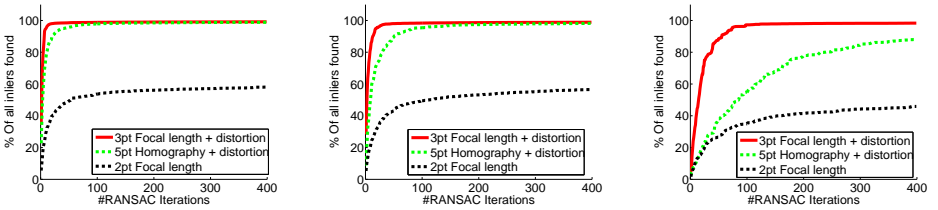
Figure 3: Number of inliers found as a function of the number of RANSAC iterations for different percentages of outliers. From left to right, the algorithm has been run on examples with 10%, 25% and 50% outliers taken from the *City* data set. In all cases the RANSAC algorithm was run 100 times and mean values were calculated. As can be seen, for moderate to large numbers of outliers, the minimal solver is superior to the overdetermined solver for homography and distortion. In neither case is the two-point solver for focal length competitive. This is expected since the two-point solver assumes zero distortion.

point DLT algorithm for estimating a homography and hence we omit a comparison with the DLT. We fixed the threshold for outlier rejection to 3 pixels and ran each algorithm in turn for 400 RANSAC iterations, keeping track of the largest inlier set found so far. We repeated this 100 times on noisy point matches from the *City* data set and computed averages. To study the influence of varying degrees of outlier contamination we also repeated the whole process for cases with 10%, 25% and 50% outliers. The results of this experiment are shown in Figure 3. As can be seen, the two-point algorithm is not competitive on this sequence and recovers half as many or fewer inliers compared to our algorithm in all cases. The behavior of the five-point algorithm is more interesting. For the case with very few inliers its performance in terms of inliers is virtually as good as for the minimal algorithm. This is because the quality of the inlier point matches is quite high in terms of pixel accuracy, which means that as long as we find a set of good quality inliers we are well served by either algorithm. For the case with 25% outlier rate we already observe a significant difference and for outlier rates of 50% and more our algorithm is clearly superior. The running time for the five point algorithm is slightly lower at around 10 milliseconds/instance in our implementation compared to 13 milliseconds for the three point algorithm. With a moderate degree of outliers in the process, this speed gain is easily eaten up by the extra RANSAC iterations required.

Although the two-point algorithm recovers less inliers than *e.g.* the three point algorithm, it still finds a substantial number of correct matches. However, the problem is that these matches are exactly the matches which agree with the assumption of zero distortion. In Figure 4 one can see the qualitative difference between correspondences produced by the three point method (with distortion) and the two-point method (no distortion). Whereas the three point solver produces matches well spread out over the images, the two-point solver recovers points grouped together near the centers of the images where the projection is reasonably well approximated by a pin-hole camera. This is problematic for two reasons, (i) the initial estimate of camera parameters will be poor and (ii) points located closely together makes for poor conditioning of the bundle adjustment step.

It should be mentioned that there are other possible ways around this problem. One could *e.g.* set an artificially high threshold of outlier rejection hoping to recover more inliers, but at the same time increasing the risk of accepting a false match as an inlier. One could also look for inliers in multiple passes by alternating bundle adjustment and inlier selection, but this process is more costly and prone to ending up in local minima. For comparison we
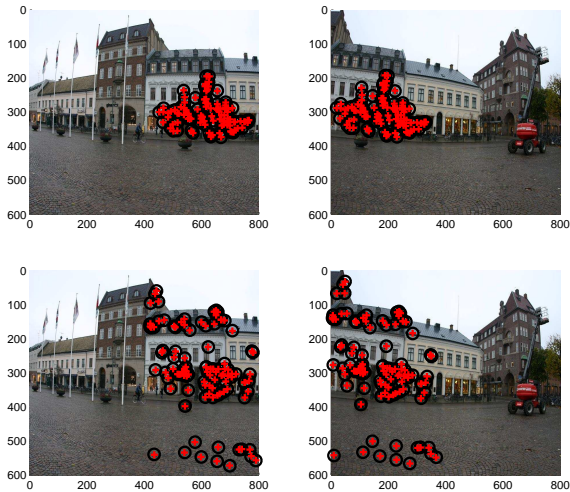
Figure 4: Point matches generated using the 2-point algorithm (top), versus our new 3-point algorithm including radial distortion (bottom). Note that the 2-point algorithm is only able to find matches in the central, undistorted portion of the images whereas the 3-point algorithm finds matches all the way to the image edge. This allows for a much more robust image alignment procedure in the presence of radial distortion.

ran the Autostitch software by Brown and Lowe [4] on the *City*, and *University* data sets. Despite not explicitly accounting for radial distortion, Autostitch was actually able to stitch together both sequences. However the final result contains visible artifacts which using the system presented in this paper we were able to avoid. Close-ups of two examples are shown in Figure 5.

6   Conclusions

We have presented a solution to the problem of estimating rotation, focal length and radial distortion from two images of the same scene undergoing pure rotation using the minimal setup with three point correspondences. The main contribution is that compared to a previous method for this problem, we are able to guarantee that the correct solution is found for all cases. Moreover, we have shown that including radial distortion at the RANSAC stage is beneficial compared to distortion free approaches in terms of number of inliers found and overall precision. An advantage of our algorithm is the ability to recover inliers evenly over the whole image where an algorithm which does not model distortion will only keep point matches close to the centers of the images. Having point matches in the center as well as close to the edges improves recognition performance as well as stability in the subsequent bundle adjustment stage. Compared to a non-minimal algorithm, we in particular do much better at higher outlier rates since a smaller correspondence set yields a smaller risk of hitting an outlier or a poor quality match.

Finally, we have investigated the practical value of the algorithm on realistic data sets and demonstrated qualitative improvements in the end result compared to a recently published stitching system.

Figure 5: Top row: Close-ups on two mistakes made by Autostitch on the sequence *City*. Bottom row: Results obtained using the system presented in this paper.

(a) *City*

(b) *University*

(c) *Canal*

Figure 6: 360 degree panoramic stitching of the sequences *City*, *University* and *Canal* using the system described in this paper. The first two sequences were shot using a fish-eye lens while the last sequence was shot with a normal lens. The stitching pipeline includes the following steps: A RANSAC stage where good point matches are established and an initial guess for geometry and calibration is estimated, a pair wise bundle adjustment step to polish the initial estimate, a global bundle adjustment step to further refine internal and external calibration parameters and finally a rendering step with basic blending.

References

[1] J. Barreto and K. Daniilidis. Fundamental matrix for cameras with radial distortion. In *IEEE International Conference on Computer Vision*, Beijing, China, 2005.

[2] D. Bayer and M. Stillman. Macaulay. www.math.columbia.edu/˜bayer/Macaulay/, 1994. URL www.math.columbia.edu/~bayer/Macaulay/. An open source computer algebra software.

[3] M. Brown, R. Hartley, and D. Nister. Minimal solutions for panoramic stitching. In *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR07)*, Minneapolis, June 2007.

[4] Matthew Brown and David G. Lowe. Automatic panoramic image stitching using invariant features. *Int. J. Comput. Vision*, 74(1):59–73, 2007. ISSN 0920-5691.

[5] M. Byröd. Fast and stable polynomial equation solving and its application to computer vision. Licentiate thesis, Centre for Mathematical Sciences LTH, Lund University, Sweden, 2008.

[6] M. Byröd, K. Josephson, and K. Åström. Improving numerical accuracy of gröbner basis polynomial equation solvers. In *Proc. 11th Int. Conf. on Computer Vision*, Rio de Janeiro, Brazil, 2007.

[7] Martin Byröd, Klas Josephson, and Kalle Åström. A column-pivoting based strategy for monomial ordering in numerical gröbner basis calculations. In *The 10th European Conference on Computer Vision*, 2008.

[8] Martin Byröd, Zuzana Kukelova, Klas Josephson, Tomas Pajdla, and Kalle Åström. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. In *Proc. Conf. Computer Vision and Pattern Recognition, Anchorage, USA*, 2008.

[9] D. Cox, J. Little, and D. O'Shea. *Ideals, Varieties, and Algorithms*. Springer, 2007. ISBN 0-387-35650-9.

[10] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6): 381–95, 1981.

[11] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proceedings of Computer Vision and Pattern Recognition Conference (CPVR)*, pages 125–132, 2001.

[12] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004. Second Edition.

[13] Hailin Jin. A three-point minimal solution for panoramic stitching with lens distortion. *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008. ISSN 1063-6919.

[14] Z. Kukelova, M. Bujnak, and T. Pajdla. Automatic generator of minimal problem solvers. In *Proc. 10th European Conf. on Computer Vision, Marseille, France*, 2008.

[15] Z. Kukelova and T. Pajdla. A minimal solution to the autocalibration of radial distortion. In *CVPR*, 2007.

[16] Z. Kukelova and T. Pajdla. Two minimal problems for cameras with radial distortion. In *Proceedings of The Seventh Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras (OMNIVIS)*, 2007.

[17] H. Li and R. Hartley. A non-iterative method for lens distortion correction from point matches. In *Workshop on Omnidirectional Vision*, Beijing China, October 2005.

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 2004.

[19] H. Stewénius. *Gröbner Basis Methods for Minimal Problems in Computer Vision*. PhD thesis, Lund University, April 2005.

[20] H. Stewénius, F. Schaffalitzky, and D. Nistér. How hard is three-view triangulation really? In *Proc. Int. Conf. on Computer Vision*, pages 686–693, Beijing, China, 2005.

[21] R.S. Szeliski. *Image Alignment and Stitching*. World Scientific, 2007.