Two View Geometry Estimation with Outliers

Olof Enqvist
www.maths.lth.se/vision/people/olofe

Fredrik Kahl
www.maths.lth.se/vision/people/fredrik

Centre for Mathematical Sciences
Lund University
Sweden

Abstract

We study the relative orientation problem for two calibrated cameras with outliers from the feature matching. In recent years there has been a growing interest in optimal algorithms for computer vision. Most people agree that to get accurate solutions to multiview geometry problems, an appropriate norm of the reprojection errors should be minimized. To this end local as well as global optimization methods have been employed.

To handle outliers though, heuristic methods still dominate the field. In this paper we address the problem of estimating relative orientation from uncertain feature correspondences. We formulate this task as an optimization problem and propose a branch-and-bound algorithm to find the optimal set of correspondences as well as the optimal relative orientation. The approach is based on geometric constraints for pairs of correspondences. The experimental results are promising, especially for omnidirectional cameras. An implementation of the algorithm is also made publicly available to facilitate further research.

1 Introduction

Geometric problems are central in computer vision. It can be the estimation of camera pose from a single image or 3D reconstruction of a whole scene from multiple images. Typically, feature points are first extracted from the images. These are then matched to corresponding points in the model or in the other images. Next these corresponding points are used to estimate the geometry. Given correct correspondences, accurate methods to estimate the geometry often exist, see [6] for a survey. However, automatic matching is difficult and avoiding erroneous matches is often impossible. This gives rise to outliers in the data, often completely ruining the accuracy of the solution.

Estimation of relative orientation is a classical problem in vision. One of the most well-known methods is the eight-point algorithm introduced by Longuet-Higgins in 1981 [10], and modified by Hartley in 1997 [8] to include normalization. Although normalization made the algorithm more robust to measurement noise, there are still algorithmic degeneracies and the algorithm breaks down in the presence of outliers. For calibrated cameras, the first guaranteed optimal algorithm based on minimizing reprojection errors appeared just recently in [7]. A severe limitation of this method is that it cannot deal with outliers among the feature correspondences.

Solving the relative orientation problem is difficult, since it is by nature non-convex and known to be plagued by local minima and ambiguous solutions [1, 12]. To handle outliers

in the correspondence set, heuristic methods like hypothesize-and-test approaches are still dominating, for example, RANSAC [5, 9]. In this paper we address the problem of uncertain feature correspondences by formulating it as a mathematical optimization problem. We provide both necessary and sufficient geometric constraints for an optimal solution. Based on this analysis, we propose an algorithm to find the optimal set of correspondences as well as the optimal relative orientation. A similar approach was proposed in [3] for the camera pose problem.

2   Preliminaries

The algorithm presented here addresses exactly the same problem as is commonly done with RANSAC, that is, trying to find the largest consistent set of correspondences. Unlike RANSAC though, the new method is guaranteed to find the globally optimal solution. In this section we introduce the constraints used in the algorithm and prove some results regarding sufficiency.

Working with calibrated cameras it is natural to represent image points using unit 3D vectors x_i defined by the projection

$$\lambda x_i = R(X_i - t), \tag{1}$$

where X_i is the 3D point, t the camera centre and R the rotation of the camera relative the world coordinate system. This way, the image plane of the ordinary perspective camera model is replaced by an image sphere. One advantage is that omnidirectional cameras can be handled in a natural way. Considering two cameras, we can assume that the first camera is placed at the origin and oriented such that the rotation is equal to the identity, and thus the relative orientation can be parameterized by a single rotation R and a translation vector t.

Definition 1. A *correspondence* is a pair of image points $(x, \bar{x})$ with x belonging to the first image and $\bar{x}$ to the second suggesting that x and $\bar{x}$ are the projections of a single 3D point.

Definition 2. Given an error tolerance ε, a correspondence $(x, \bar{x})$ is said to be *consistent* with a relative orientation (R, t), if there exists a 3D point X such that its angular reprojection errors satisfy

$$\angle(x, X) < \varepsilon \quad \text{and} \quad \angle(\bar{x}, R(X - t)) < \varepsilon. \tag{2}$$

We can now state the optimization problem that we are aiming to solve:

Problem 1. Given two sets of image points $\{x_i\}$ and $\{\bar{x}_j\}$ with hypothetical correspondences $(x_k, \bar{x}_k)$, $k = 1, \ldots, N$ and a prescribed error threshold ε, compute the relative orientation of the cameras which is consistent with as many correspondences as possible.

Our aim is an algorithm that is guaranteed to solve this optimization problem. To this end, we first examine the geometrical constraints induced by a set of correspondences.

Consider feature points from two calibrated images and a set of hypothetical correspondences. Suppose that we know the relative orientation of these images and want to determine which correspondences are correct. We introduce new bases in both cameras with coinciding z-axes parallel to the epipole of camera 1 and define α to be the relative rotation angle of these bases (see Figure 2).

Using these bases, consider the spherical coordinates of a projected 3D point in the first camera $(\theta_{proj}, \varphi_{proj})$ and in the second camera $(\bar{\theta}_{proj}, \bar{\varphi}_{proj})$, where the first coordinate is the
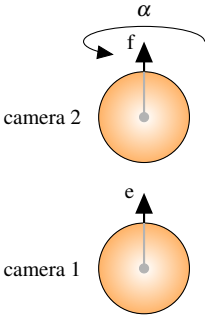
Figure 1: The z-axis goes through the two camera centres and hence the epipole e is parallel with the z-axis. Similarly, the antipodal epipole f is parallel to the z-axis.
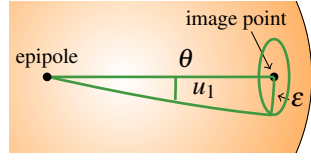

Figure 2: To compute the uncertainty $u = u_1 + u_2$ when $\bar{\theta} > \theta$, we use the spherical law of sines, see [13].
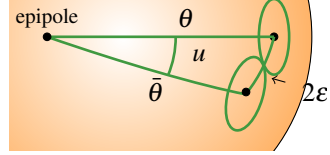

Figure 3: Here we use the spherical law of cosines, see [13]. The figure shows the two image spheres superposed. The uncertainty u is the azimuthal angle between the image points such that the ε-circles still intersect.

polar angle and the second the azimuthal angle. It must hold that $\bar{\theta}_{proj} > \theta_{proj}$ and $\varphi_{proj} = \bar{\varphi}_{proj} + \alpha$. The first constraint just reflects the fact that 3D points should be in front of both cameras and the second that the image points should be coplanar with the camera centres. Since we are interested in points with reprojection error less than ε, similar constraints hold for the *measured* image points. The constraint on the polar angle of corresponding image points is the simplest,

$$\bar{\theta} + \varepsilon > \theta - \varepsilon. \tag{3}$$

For the azimuthal angle, the error tolerance ε gives rise to a larger angular uncertainty u. Figure 2 shows how to compute this if $\bar{\theta} > \theta$. We get

$$u = \arcsin\left(\sin\varepsilon / \sin\theta\right) + \arcsin\left(\sin\varepsilon / \sin\bar{\theta}\right) \tag{4}$$

If $\bar{\theta} < \theta$, we get a different formula. Figure 2 shows how to compute the uncertainty in this case, given that (3) is still satisfied. We get

$$u = \arccos\left(\left(\cos\left(2\varepsilon\right) - \cos\theta\cos\bar{\theta}\right) / \left(\sin\theta\sin\bar{\theta}\right)\right). \tag{5}$$

For a consistent correspondence, the difference between the azimuthal angles measured in the first and second image, that is $\varphi - \bar{\varphi}$, satisfy

$$|\varphi - \bar{\varphi} - \alpha| < u. \tag{6}$$

Theorem 1. *Let $(x,\bar{x})$ be a feature correspondence with spherical coordinates (θ,φ) and $(\bar{\theta},\bar{\varphi})$, respectively. If for some α, both constraints (3) and (6) are satisfied then $(x,\bar{x})$ is consistent with the given relative orientation.*

Proof. It is clear from (6) that we can pick coplanar vectors x' and $\bar{x}'$ such that $\angle(x,x') < \varepsilon$ and $\angle(\bar{x},\bar{x}') < \varepsilon$. Furthermore, (3) guarantees that we can pick these such that they intersect with positive depths. This intersection is the desired 3D point of Definition 2. $\qquad\square$

The idea is to parameterize the relative orientation by specifying these coordinate systems in the original coordinates of the cameras. More precisely, let e be the epipole and thus the new z axis of camera 1 given in the old camera coordinate system and f the same axis but given in the camera coordinate system of camera 2. Knowing e allows us to compute the polar angles θ and the azimuthal angles φ for the first image, and similarly, knowing f in the second image allows us to compute $\bar{\theta}$ and $\bar{\varphi}$.

We eliminate the angle α by considering pairs of correspondences instead. Using (6) for a pair of correspondences yields

$$|(\varphi_j - \bar{\varphi}_j) - (\varphi_k - \bar{\varphi}_k)| < u_j + u_k. \qquad (7)$$

Another way to view these constraints is by considering the angle γ_{jk} between two epipolar planes (see Figure 4). What (7) states is that apart from the uncertainty $u_j + u_k$, this angle should not depend on whether we measure it in the first camera, $\varphi_j - \varphi_k$, or in the second, $\bar{\varphi}_j - \bar{\varphi}_k$.
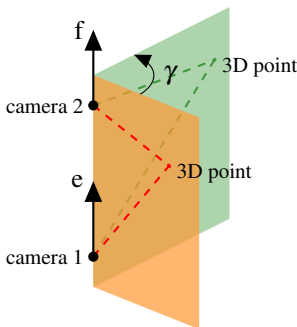

Figure 4: The angle γ between two epipolar planes.

By construction, these constraints are necessary. Let us investigate the question of sufficiency. We need the following lemma.

Lemma 1. *Consider a set of intervals I_k on the unit circle such that $|I_k| < 2\pi/3$ for all k. If the intersection $I_j \cap I_k$ is non-empty for any pair (j,k), then $\bigcap_k I_k$ is non-empty as well.*

Proof. Pick any interval I_k and let m be the centre of this interval. Note that the distance between I_k and the point $m + \pi$ is larger than $2\pi/3$ and thus $m + \pi$ lies in no interval. Thus we can cut the unit circle at $m + \pi$ and map it to the real line $\mathbb{R}$. The theorem now follows from Helly's theorem [2]. $\qquad \square$

Now, if all pairwise correspondences fulfill (7) with uncertainty intervals less than $2\pi/3$, then from the above lemma we know that there exists an angle α such that constraint (6) is fulfilled for each single correspondence. Applying Theorem 1 proves sufficiency. To conclude, we have shown the following result.

Theorem 2. *Let $(x_j, \bar{x}_j)$, $j = 1, \ldots, N$ be image correspondences with spherical coordinates (θ_j, φ_j) with respect to epipole e in the first image and $(\bar{\theta}_j, \bar{\varphi}_j)$ with respect to antipodal epipole f in the second image. If all correspondences satisfy the positive depth constraint in (3) and all pairs of correspondences satisfy constraint (7) with uncertainties $u_j + u_k$ less than $2\pi/3$, then there exists a relative orientation (R, t) to which all correspondences are consistent.*

Remark 1. The extra hypothesis that the uncertainty intervals should have length at most $2\pi/3$ is just an annoying technicality. Larger intervals would generally correspond to image points very close to the epipole (within angles $\varepsilon < \theta < 2\varepsilon$), and even if such points exist it is very likely that the conclusion of the lemma still holds. To get around this problem one could split long intervals and still get guaranteed convergence of our algorithm (see next section), but we have not found any problem instance where this is necessary in practice.

2.1 Handling Outliers

How can these constraints be used to handle outliers? Regarding the constraint on the polar angles θ it is simple enough. A correspondence which does not satisfy this constraint must be an outlier. For the pairwise constraints on φ it is more complicated. A violated constraint only implies that one of the involved correspondences is incorrect, but how do we determine which one? It may be that they are both involved in a number of violated constraints.

It turns out that the answer can be found in graph theory. Consider a graph with one vertex for each correspondence. Remove correspondences which do not satisfy the constraint on the polar angles. Then check the pairwise constraints on the azimuthal angles and mark any inconsistence by adding an edge between the inconsistent correspondences. Solving Problem 1 means removing as few correspondences as possible while explaining all inconsistencies, or equivalently, removing as few vertices as possible while covering all edges. This is known as the vertex cover problem and is a classic graph theoretic problem. We will not discuss it in detail here but note that it is an NP-hard problem and that it can always be solved for example using a branch and bound. A more thorough discussion on using pairwise constraints to remove outliers can be found in [4].

For our purposes very simple approximations for this problem will be sufficient. Assume we are looking for a solution with less than N outliers. Then any correspondence involved in more than N violated constraints can be removed. Furthermore, we can estimate number of outliers for a given relative orientation by using the well known factor-2 approximation for vertex cover (see [14]).

3 Algorithm

The previous section provided us with necessary and sufficient geometrical constraints on the relative orientation. In this section we show how to evaluate these constraints and propose a branch and bound scheme to find the optimal solution. Recall that according to Theorem 2, one can solve Problem 1 by finding the epipole e in the first image and the antipodal epipole f in the second image, which yields the largest set of pairwise consistent correspondences.

The possible values for these epipole parameters lie in $\mathbb{S}^2 \times \mathbb{S}^2$ so we need some method to search this space. We will use a branch and bound approach dividing the spheres into smaller and smaller spherical triangles (see Figure 5). To evaluate the pairwise constraints in (7), we need bounds on the possible values of the azimuthal angles. It is sufficient to consider $\gamma_{jk} = \varphi_j - \varphi_k$ which is the angle between two epipolar planes as illustrated in Figure 4. Hence, our next goal is to derive upper and lower bounds on the epipolar plane angle γ_{jk} for a pair of image points given that the epipole e lies in a spherical triangle. Similarly, we need bounds on $\bar{\gamma}_{jk}$ given that the antipodal epipole f lies in another spherical triangle. Our first observation is that it is sufficient to consider the boundaries of the spherical triangles.
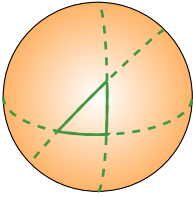
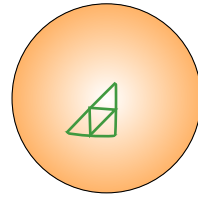Figure 5: A spherical triangle is a triangular area on the sphere, bounded by great circles.

Figure 6: In the division step a triangle is divided into four smaller ones by dividing its sides at their midpoints.

Theorem 3. *Given two image points x_j and x_k and a spherical triangle of possible epipoles. If either $\pm x_j$, $\pm x_k$ lie in the triangle then γ_{jk} can assume any value. Otherwise the maximal and minimal values for γ_{jk} are obtained on the triangle boundary.*

To prove this claim we will use an explicit formula for the epipolar angle γ (we temporarily drop the subindex to simplify notations). We choose a special coordinate system such that the two image points have coordinates $(0, \pm a, b)$ and let $e = (e_x, e_y, e_z)$ be the position of the epipole and assume $e_x \geq 0$ implying $\gamma \geq 0$. Let us first compute the normals of the two epipolar planes,

$$e \times x_j = \begin{pmatrix} be_y - ae_z \\ -be_x \\ ae_x \end{pmatrix}, \quad e \times x_k = \begin{pmatrix} be_y + ae_z \\ -be_x \\ -ae_x \end{pmatrix}. \tag{8}$$

Then consider their scalar product,

$$b^2 e_y^2 - a^2 e_z^2 + b^2 e_x^2 - a^2 e_x^2 = \cos\gamma \left(\sqrt{(be_y - ae_z)^2 + e_x^2} \sqrt{(be_y + ae_z)^2 + e_x^2} \right). \tag{9}$$

The fact that $e_x^2 + e_y^2 + e_z^2 = 1$ and some simple calculus yield

$$e_y^2 - a^2 + b^2 e_x^2 = 2abe_x \cot\gamma. \tag{10}$$

We will only prove the case $e_x > 0$ here. For $e_x < 0$ an almost identical proof can be used and tying them together is a simple question of going through some different cases.

Proof. $(e_x > 0)$ We will show that γ as a function of the epipole has no local extrema and hence minimal and maximal values of γ are attained at the boundary. A good start would be to prove that

$$g(e_x, e_y) = 2ab \cot\gamma = \frac{e_y^2 - a^2 + b^2 e_x^2}{e_x} \tag{11}$$

has no stationary points. Note that $g'_y = 0$ only if $e_y = 0$ and $g'_x = 0$ yields

$$0 = \frac{2b^2 e_x^2 - e_y^2 + a^2 - b^2 e_x^2}{e_x^2} = \frac{b^2 e_x^2 + a^2}{e_x^2}, \tag{12}$$

which is never satisfied. Thus it remains to examine the special cases. For $e_x^2 + e_y^2 = 1$ we get

$$h(e_x) = 2ab \cot\gamma = \frac{1 - e_x^2 - a^2 + b^2 e_x^2}{e_x} = \frac{b^2 - a^2 e_x^2}{e_x} = \frac{b^2}{e_x} - a^2 e_x \tag{13}$$

$$h'(e_x) = -\frac{b^2}{e_x^2} - a^2 \tag{14}$$

which is never zero. Finally the end point $e_x = 1$ turns out to be a saddle point and not an extremum. □

To check the constraints on the epipolar plane angles γ_{jk}, we need to compute upper and lower bounds for this angle on parts of great circles corresponding to the boundary of a spherical triangle. We have experimented with several techniques to do this, mainly by setting up the problem as a system of polynomial equations. The most effective turned out to be a simple representation by angles.

Consider a great circle parameterized by a single angle ξ. Let d_i be the angular distance from image point i to this great circle. One way to compute γ_{ij} is to compute the angle relative to the great circle for each image points (see Figure 7) and then compute pairwise differences. In spherical coordinates this can be seen as computing φ_j

$$\varphi_j(\xi) = \pm \operatorname{arccot}(\sin\xi \cot d_j). \tag{15}$$

where the formula is obtained using the spherical law of cosines and the spherical law of sines on the right-angled triangle in Figure 7. The sign in (15) depends on which side of the great circle the image points lies. The angle between two epipolar planes thus gets the form

$$\gamma_{jk}(\xi) = \pm \operatorname{arccot}(\sin\xi \cot d_j) \pm \operatorname{arccot}(\sin(\xi + \beta_k)\cot d_k). \tag{16}$$

There are two difficulties in approach. First of all, finding the maximum and minimum values of (16), generally requires solving a sixth degree polynomial equation. Furthermore, the number of functions to optimize is quadratic in the number of correspondences (since we are considering all pairs). To address both these problems we compute simple under- and overestimators for φ_i. These can then be used to produce under- and overestimators for the γ_{jk}'s in the first image (and similarly for $\bar{\gamma}_{jk}$ in the second image). This avoids the difficult optimization of γ_{jk} and most work is performed on the φ_j's and is thus linear in the number of image points.

Simple under- and overestimators for a C^3 function $\varphi(\xi)$ follows from the taylor expansion.

$$\varphi(\xi) \geq \varphi(\xi_0) + (\xi - \xi_0)\varphi'(\xi_0) + \frac{(\xi - \xi_0)^2}{2}\varphi''_{min}$$
$$\varphi(\xi) \leq \varphi(\xi_0) + (\xi - \xi_0)\varphi'(\xi_0) + \frac{(\xi - \xi_0)^2}{2}\varphi''_{max} \tag{17}$$

This is the type of estimations that we will use. In many cases these will show that γ_{jk} is in fact monotonic on the interval in question. In these cases it is sufficient to consider the endpoints (i.e. corners of the spherical triangle) and computation is very fast. Otherwise we use the upper and lower bounds given by the estimators. Note that since we are using a first order approximation, the approximation error decreases quadratically with the radius of the triangles.

Let us quickly go through the different steps of the algorithm. To initiate the branch and bound search we divide the search space $\mathbb{S}^2 \times \mathbb{S}^2$ into 64 starting blocks. A block here is a pair of spherical triangles indicating the position of the epipole e of camera 1 and the
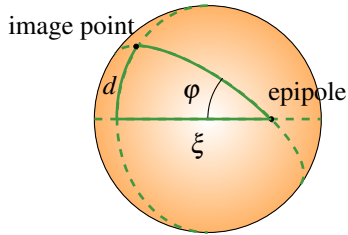
Figure 7: We first compute angles to the great circle and then take pairwise differences to yield the epipolar plane angles $\gamma_{ij} = \varphi_i - \varphi_j$.

antipodal epipole f of camera 2. We also need a bound on the optimal solution. This is a starting guess for the number of outliers of the optimal solution and will be updated as the algorithm progresses. If the starting guess is to low the algorithm is restarted with a higher value.

Algorithm 1.

Iterate until desired precision is reached:
*   1. Pick a box from the queue.*
*   2. Try to detect and remove outliers.*
*   3. Try to discard the box.*
*   4. If the box cannot be discarded:*
*   - Divide the box and update the queue.*
*   - Try improve the bound on the optimum.*
*   5. Remove the box from the queue.*

Finally a word on updating the bound on the optimum, that is, finding better and better solutions. We do this by considering the centres of those boxes that could not be discarded. For these points we compute all the constraints and count the number of outliers. As the boxes get smaller and smaller, the optimal number of outliers will eventually be found.

4   Experiments

The algorithm was implemented in C and tried out on a 3.0 GHz Intel DualCore with 3 GB RAM. We kept track of lower and upper bounds on the optimal solution (in terms of the number of outliers) and stopped execution when the gap between the upper and lower bound was smaller than 3.

4.1   Valbonne Church

The algorithm was evaluated on images from the Valbonne Church data set[1]. From the 15 images all pairs were formed. SIFT features were extracted and matched. The matching criterion of [11] was used with a threshold of 0.6.

The threshold was set to 0.0005 radians and the starting guess for the number of outliers to 5. The algorithm was terminated if the gap between lower and upper bound was less than 3 or if the average uncertainty of the epipole was less than 2 degrees.

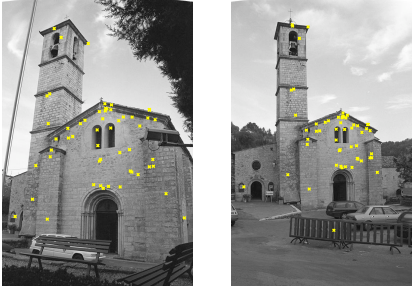[1]The Valbonne data set was provided by ROBOTVIS, INRIA.

Figure 8: Example images from the Valbonne Church data set.
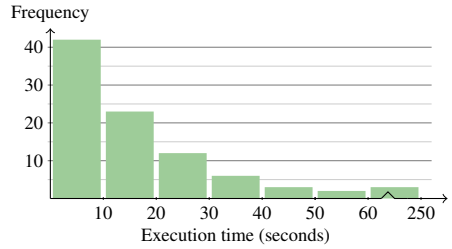

Figure 9: Execution times for the Valbonne data set. Between 10 and 100 points in each experiment (on average 64). Around 5% outliers.

4.2 Ladybug Sequence

We also tested on the omnidirectional images used in [7]. Among other things, execution times depended on the distance between the cameras. This is natural since estimating the epipole - being the direction of translation - is difficult when the length of the translation is very small.

For these experiments the error threshold was set to 0.002 radians and the starting guess for the number of outliers was initiated to 5. If there was no solution for the current threshold was increased with 5 until a solution could be found.
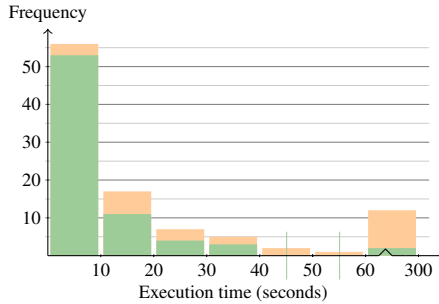

Figure 10: Execution times for the ladybug data set. Between 20 and 100 points in each experiment (on average 64). Around 10% outliers. Green indicates camera pairs with a greater base line and orange the more difficult short base line cases.

5 Conclusion

To our knowledge, the proposed algorithm is the first that is guaranteed to find the optimal relative orientation in the presence of outliers. Theoretic discussions as well as experiments demonstrate its ability to compute the globally optimal solution while discarding outliers. Execution times are still high for real-time applications but we believe that a significant speedup is possible by saving results that can be re-used and by working on the function approximations.

References

[1] A. Chiuso, R. Brockett, and S. Soatto. Optimal structure from motion: local ambiguities and global estimates. *Int. Journal Computer Vision*, 39(3):195–228, 2000.

[2] J. Eckhoff. Helly, radon and carathéodory type theorems. In *Handbook of Convex Geometry*, pages 389–448. North-Holland, Amsterdam, 1993.

[3] O. Enqvist and F. Kahl. Robust optimal pose estimation. In *European Conf. Computer Vision*, pages 141–153, Marseille, France, 2008.

[4] O. Enqvist, K. Josephson, and F. Kahl. Optimal correspondences from pairwise constraints. In *Int. Conf. Computer Vision*, Kyoto, Japan, 2009.

[5] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:381–395, 1981.

[6] R. Hartley and F. Kahl. Optimal algorithms in multiview geometry. In *Asian Conf. Computer Vision*, Tokyo, Japan, 2007.

[7] R. Hartley and F. Kahl. Global optimization through rotation space search. *Int. Journal Computer Vision*, 82(1):64–79, 2009.

[8] R. I. Hartley. In defense of the eight-point algorithm. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.

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

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

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

[12] R. Szeliski and S. B. Kang. Shape ambiguities in structure from motion. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(5), May 1997.

[13] I. Todhunter. *Spherical Trigonometry: For the Use of Colleges and Schools*. MacMillan and Co., 1886.

[14] V. Vazirani. *Approximation Algorithms*. Springer, 2001.