

# Resultant Based Incremental Recovery of Camera Pose from Pairwise Matches

Yoni Kasten

Meirav Galun

Ronen Basri

Weizmann Institute of Science

{yoni.kasten,meirav.galun,ronen.basri}@weizmann.ac.il

## Abstract

*Incremental (online) structure from motion pipelines seek to recover the camera matrix associated with an image  $I_n$  given  $n - 1$  images,  $I_1, \dots, I_{n-1}$ , whose camera matrices have already been recovered. In this paper, we introduce a novel solution to the six-point online algorithm to recover the exterior parameters associated with  $I_n$ . Our algorithm uses just six corresponding pairs of 2D points, extracted each from  $I_n$  and from any of the preceding  $n - 1$  images, allowing the recovery of the full six degrees of freedom of the  $n$ 'th camera, and unlike common methods, does not require tracking feature points in three or more images. Our novel solution is based on constructing a Dixon resultant, yielding a solution method that is both efficient and accurate compared to existing solutions. We further use Bernstein's theorem to prove a tight bound on the number of complex solutions. Our experiments demonstrate the utility of our approach.*

## 1. Introduction

This paper addresses the problem of recovering camera position and orientation in a new image  $I_n$  given a stream of  $n - 1$  images  $I_1, \dots, I_{n-1}$  whose parameters are known. Efficient and accurate solutions to this problem are important in a variety of online applications such as online robot positioning, incremental structure recovery from video images, and streaming applications.

Incremental structure from motion (SFM) pipelines [29, 2, 36] commonly compute the next camera pose by matching triangulated points (i.e., points whose depths have been recovered) from  $I_1, \dots, I_{n-1}$  to image points from  $I_n$ , e.g., using PnP algorithms [3, 12, 23, 7]. This process requires tracking feature points in three or more images. Alternative methods compute this pose by integrating information from two (or more) essential matrices (e.g., [16]). This approach utilizes an excess number of matching pairs – at least 10 pairs are needed to recover two essential matrices. Recent work [30, 37] showed that six pairs of points suffice to determine the next camera pose. Such matching pairs can relate

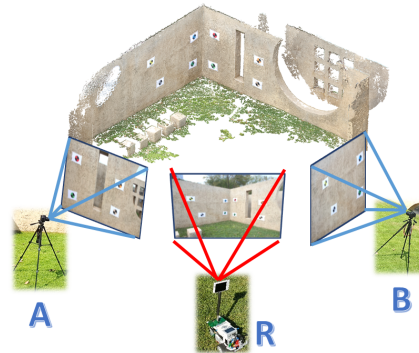


Figure 1: Images of an outdoor scene containing an *internal* wall corner with eight landmarks. Stationary cameras A and B, whose parameters are known, produce non-overlapping images. A robot mounted camera (R), whose parameters are unknown, is moving in the scene. Our method is used to recover the robot position and orientation from R-A and R-B matching pairs.

the new image  $I_n$  to *any* of the previous images  $I_1, \dots, I_{n-1}$  (in particular each matching point may relate  $I_n$  to a different image).

Figure 1 illustrates this setup. The figure shows a scene with eight landmarks placed on two walls pictured by three cameras. Each of the (fully calibrated) cameras A and B is positioned such that it sees only four of the landmarks on one of the walls. The third, robot mounted camera, whose pose is unknown, sees all the landmarks. Our goal is to recover its pose. Note that none of the points can be triangulated because the fields of view of the two calibrated cameras are non-overlapping. Also, it is not possible to recover the essential matrices that involve the third camera since they share only four landmarks with cameras A and B, and moreover these landmarks are coplanar. Nevertheless, with six pairs of matching landmarks it is possible to recover the pose of the third camera.

Below we introduce a novel solution to the online, 6-point algorithm to recover the position and orientation of a new camera in sequential, multiview SFM. We use quaternions to obtain a succinct polynomial system in the un-

known pose parameters. We compute the mixed volume of the associated Newton polytope to prove analytically that the system gives rise to 64 complex solutions and show empirically that it typically produces roughly 23 real solutions. This rigorous argument confirms previous observations based on random assignments of coefficients [30]. We next symbolically derive a solution method by constructing a Dixon resultant, which we implement efficiently. Our numerical experiments indicate that our methods produces significantly more accurate solutions than efficient, Gröbner base solutions [30, 37]. ([6]’s solutions are nearly as accurate as ours, but this method is significantly slower.) We further demonstrate the utility of our approach in sequential SFM pipelines.

### 1.1. Related work

Stewenius et al. [30] introduced an approach for recovering the relative pose of “generalized cameras”, where a generalized camera can have multiple focal centers whose relative positions are known. Our problem is a special case of their setting, since the known  $n - 1$  cameras can be considered a single generalized camera. They further derived a solution for their system of polynomial equations (which is different from our system) by constructing a Gröbner basis. They then used this basis to construct an action matrix whose eigen-system returns the 64 complex solutions of the problem. Later work by Larsson et al. [21, 22], inspired by [19], introduced an automatic generator of action matrices, which they applied to Stewenius et al.’s formulation.

Both solvers [30, 21] approach the problem by formulating an expanded set of polynomials  $C\mathbf{x} = 0$ , where  $C$  is called an elimination template matrix and  $\mathbf{x}$  is the vector of all monomials occurring in these polynomials. The action matrix is the bottom-right sub-matrix of  $C$  obtained after applying Gauss-Jordan elimination to  $C$ . This elimination process may occasionally produce a numerically unstable action matrix, resulting in numerical inaccuracies, large residuals, and subsequently large pose errors. Byrod et al. [6] attempted to rectify these numerical instabilities by first constructing yet a larger template matrix with more polynomials and additional monomials. They replaced elimination by the more stable  $QR$  decomposition with column pivoting to produce a stable action matrix. Their method however, is problem instance dependent, and so their implementation is considerably slower than previous methods – it is in fact as slow as iterative methods such as homotopy continuation.

The solvers, mentioned above, can handle most six-point configurations, but, as noted in [37], they may become degenerate when more than three points come from a single known camera. Zheng and Wu [37] addressed this special case. Their approach is related to [30] and is subject to similar inaccuracies. Also related are generalized PnP algorithms [8, 32], which align a 3D object to multiple cameras.

## 2. Method

We present our method in this section. We introduce our polynomial system in Sec. 2.1, convert it to a quaternion representation (Sec. 2.2), prove an upper bound on the number of complex solutions (Sec. 2.3), and finally derive its Dixon resultant to develop an efficient solution method (Sec. 2.4).

### 2.1. Problem definition

Our input consists of point correspondences extracted from  $n$  input images,  $I_1, \dots, I_n$ , ( $n \geq 3$ ). We assume that all cameras are internally calibrated and, further, that the exterior parameters of the first  $n - 1$  cameras in a global coordinate system are known, i.e., the camera positions,  $\mathbf{t}_i \in \mathbb{R}^3$ , and orientations  $R_i \in SO(3)$  ( $1 \leq i \leq n - 1$ ), are given. Our goal is to recover the position,  $\mathbf{t}_n$ , and orientation,  $R_n$ , used to produce the new image,  $I_n$ . Below we denote our unknowns  $R = R_n$  and  $\mathbf{t} = \mathbf{t}_n$ , omitting their subscript to simplify notation.

To determine the 6 degrees of freedom in  $R$  and  $\mathbf{t}$  we use 6 pairs of corresponding points  $\{\mathbf{p}_k, \mathbf{p}'_k\}_{k=1}^6$ , where  $\mathbf{p}_k \in I_n$ ,  $\mathbf{p}'_k \in I_{i_k}$ , and  $1 \leq i_k \leq n - 1$ . We further assume that  $|\bigcup_{k=1}^6 \{i_k\}| \geq 2$ , i.e., that the corresponding points relate points in  $I_n$  to at least two of the first  $n - 1$  images, and that together with  $I_n$ , these images are produced by cameras that are not all collinear. (We further discuss collinear camera settings in Sec. 2.4.) Each corresponding pair is then related by the appropriate essential matrix  $\mathbf{p}_k^T E_{n,i_k} \mathbf{p}'_k = 0$ , where we denote by  $E_{n,i_k}$  the essential matrix between  $I_n$  and  $I_{i_k}$ . To simplify our notations we next omit the subscript  $k$ , so we write this relation as

$$\mathbf{p}^T E_{n,i} \mathbf{p}' = 0. \quad (1)$$

Our approach relies on expressing the essential matrix  $E_{n,i}$  in terms of the camera positions and orientations (expressed in a global coordinate frame) in  $I_n$  and  $I_i$ . Using derivation introduced in [4]:

$$E_{n,i} = R^T (T - T_i) R_i. \quad (2)$$

$T = [\mathbf{t}]_{\times}$  and  $T_i = [\mathbf{t}_i]_{\times}$  respectively are skew-symmetric matrices representing cross products with  $\mathbf{t}$  and  $\mathbf{t}_i$ . Plugging (2) into (1) and collecting known quantities we obtain

$$\mathbf{p}^T R^T (\mathbf{s} \times \mathbf{t} + \mathbf{b}) = 0, \quad (3)$$

where  $\mathbf{p}$ ,  $\mathbf{s} = -R_i \mathbf{p}'$ , and  $\mathbf{b} = -T_i R_i \mathbf{p}'$ , are known quantities, while  $R$  and  $\mathbf{t}$  are unknown. Each of the six corresponding pairs contributes one equation of the form (3). This equation is bilinear in the unknowns  $R$  and  $\mathbf{t}$  and homogeneous in  $R$  (but not in  $\mathbf{t}$ ). Restricting  $R$  to be rotation further provides a quadratic constraint of the form  $R^T R = I$ , where  $I$  denotes the identity matrix.

We note the difference between (3) and the usual essential matrix relation for two images that is solved in [14, 28, 24]. With two images we can set the global coordinates to coincide with those of  $I_i$ , obtaining  $R_i = I$  and  $\mathbf{t}_i = 0$ , which yields  $\mathbf{s} = -\mathbf{p}'$  and  $\mathbf{b} = 0$ . This makes (3) homogeneous in the translation parameters,  $\mathbf{t}$ . Consequently,  $\mathbf{t}$ , and likewise  $E_{n,i}$ , can be recovered only up to scale, requiring a mere 5 pairs of points. The use of additional images makes our equations inhomogeneous in  $\mathbf{t}$ , allowing us, when the cameras are not all collinear, to recover all the 6 degrees of freedom in the exterior parameters corresponding to  $I_n$ .

## 2.2. Quaternion representation

For our equations of the form (3), similar to [30, 14, 11], we use quaternions [13] to eliminate the orthogonality constraints,  $R^T R = I$ . A quaternion is represented by a 4-vector  $\dot{\mathbf{q}} = (r; \mathbf{v}) \in \mathbb{R}^4$ , where  $r \in \mathbb{R}$  denotes its scalar (real) part and  $\mathbf{v} \in \mathbb{R}^3$  represents its vector (imaginary) part. We use the semicolon symbol to denote column concatenation. The space of quaternions is endowed with a product operation. The Hamilton product between two quaternions,  $\dot{\mathbf{q}}_1 = (r_1; \mathbf{v}_1)$  and  $\dot{\mathbf{q}}_2 = (r_2; \mathbf{v}_2)$ , is defined as

$$\dot{\mathbf{q}}_1 \dot{\mathbf{q}}_2 = (r_1 r_2 - \mathbf{v}_1^T \mathbf{v}_2; r_1 \mathbf{v}_2 + r_2 \mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2), \quad (4)$$

where the ‘ $\times$ ’ symbol denotes the cross product between two vectors in  $\mathbb{R}^3$ . The conjugate of  $\dot{\mathbf{q}} = (r; \mathbf{v})$  is defined as  $\dot{\mathbf{q}}^* = (r; -\mathbf{v})$  and its reciprocal by  $\dot{\mathbf{q}}^{-1} = \dot{\mathbf{q}}^* / \|\dot{\mathbf{q}}\|^2$ . Rotations are represented by unit quaternions (so in particular  $\dot{\mathbf{q}}^{-1} = \dot{\mathbf{q}}^*$ ). A rotation by an angle of  $\theta$  about an axis  $\mathbf{u} \in S^2$  ( $S^2$  is the unit sphere in  $\mathbb{R}^3$ ), denoted  $R = R(\theta, \mathbf{u})$ , is represented by  $\dot{\mathbf{q}} = (\cos \frac{\theta}{2}; \sin \frac{\theta}{2} \mathbf{u})$ . Applying  $R$  to a point  $\mathbf{p} \in \mathbb{R}^3$  is expressed via conjugation as follows,

$$(0; R\mathbf{p}) = \dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*, \quad (5)$$

where we define  $\dot{\mathbf{p}} = (0; \mathbf{p})$ .

Using a quaternion formulation, (3) can be written as

$$\text{vec}(\dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*)^T (\mathbf{s} \times \mathbf{t} + \mathbf{b}) = 0, \quad (6)$$

where we denote by  $\text{vec}(\cdot)$  the vector part of a quaternion. This formulation includes 7 unknowns, the four components of  $\dot{\mathbf{q}}$  and the three components of  $\mathbf{t}$ . The former are further constrained by  $\|\dot{\mathbf{q}}\|^2 = 1$ . However, since (6) is homogeneous in  $\dot{\mathbf{q}}$  this constraint can be omitted and replaced by fixing one of the entries of  $\dot{\mathbf{q}}$ . Below we fix the real part of  $\dot{\mathbf{q}}$ , which we denote by  $q_1$ , to 1. This restricts the angle of rotation  $\theta$  by requiring  $\cos(\theta/2) \neq 0$ , i.e.,  $\theta \neq \pi$ . This case can be handled separately, e.g., by rotating the global coordinate system. In summary, our formulation has 6 unknowns, the vector part of  $\dot{\mathbf{q}}$  and the components of  $\mathbf{t}$ , and so it can be solved by providing 6 corresponding pairs of points, each supplies one polynomial equation. We note

that these polynomials are cubic; they are quadratic in  $\dot{\mathbf{q}}$  and linear in  $\mathbf{t}$ .

A further change of variables, similar to a formulation suggest by Horn [14] for essential matrices, can be applied to reduce the degree of these polynomial equations to 2. This will serve us to prove a tight bound on the number of solutions and will be useful in devising an efficient solution scheme. To define the change of variables note that the first term in (6) represents a triple product of the form  $\text{vec}(\dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*)^T (\mathbf{s} \times \mathbf{t})$ . Triple products are invariant to cyclic permutations, and so it can be replaced by  $\mathbf{s}^T (\mathbf{t} \times \text{vec}(\dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*))$ . By the properties of the Hamilton product, let  $\dot{\mathbf{t}} = (0; \mathbf{t})$ , we can write

$$\mathbf{t} \times \text{vec}(\dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*) = \text{vec}(\dot{\mathbf{t}} \dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*). \quad (7)$$

We now let  $\dot{\mathbf{d}} = \dot{\mathbf{t}} \dot{\mathbf{q}}$ , then (6) becomes

$$\mathbf{s}^T \text{vec}(\dot{\mathbf{d}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*) + \mathbf{b}^T \text{vec}(\dot{\mathbf{q}} \dot{\mathbf{p}} \dot{\mathbf{q}}^*) = 0. \quad (8)$$

This polynomial is quadratic in  $\dot{\mathbf{q}}$  and bilinear in  $\dot{\mathbf{q}}$  and  $\dot{\mathbf{d}}$ . Note that  $\dot{\mathbf{d}}$  has 4 entries, increasing the number of variables to 7. The existence of an additional constraint is therefore implied. By the definition of  $\dot{\mathbf{d}}$ , and noting that  $\dot{\mathbf{q}} \dot{\mathbf{q}}^* = 1$  we obtain that  $\dot{\mathbf{t}} = \dot{\mathbf{d}} \dot{\mathbf{q}}^*$ . The following relation ensures that the real part of  $\dot{\mathbf{t}}$  will be identically 0,

$$\dot{\mathbf{d}}^T \dot{\mathbf{q}} = 0. \quad (9)$$

We obtain in total a system of 7 polynomial equations in 7 unknowns; 6 corresponding pairs yield 6 equations of type (8), along with the additional constraint (9).

## 2.3. Number of solutions

Next, we wish to determine the number of solutions to our polynomial system of equations. Previous work [30, 21] used random coefficient assignments to argue that the problem gives rise to 64 complex solutions. Here we use Bernstein’s bound [5] to prove this rigorously, confirming these early observations.

One way to obtain a bound on the number of complex solutions is by using Bezout’s Theorem [9]. According to this theorem, a generic system of  $m$  polynomials in  $m$  variables of degree  $d$  should have up to  $d^m$  common complex roots. Applying this to (6), which consists of 6 cubic polynomials yields a bound of  $3^6 = 729$  solutions. A tighter bound is obtained using the formulation of (8) and (9). This system consists of 7 quadratic polynomials, yielding a bound of  $2^7 = 128$  complex solutions.

Bezout’s theorem allows the polynomials to include all terms up to degree  $d$  with independent coefficients. In fact, a tighter bound can be derived by noting that our polynomials are sparse, in the sense that they involve only a subset of the terms. The tighter bound is obtained by applying the Bernstein-Khovanskii-Kushnirenko (BKK) Theorem [5], which relies on a remarkable connection between

polynomials and convex polytopes. A polynomial  $f$  in  $m$  variables,  $\mathbf{x} = (x_1, \dots, x_m)$ , is a finite sum of terms of the form  $c_\alpha \mathbf{x}^\alpha$ , with  $c_\alpha \neq 0$ ,  $\alpha = (\alpha_1, \dots, \alpha_m)$ , and we use the multi-index notation  $\mathbf{x}^\alpha = \prod_{i=1}^m x_i^{\alpha_i}$ . The collection of vectors of powers for  $f$ ,  $\{\alpha\} \subset \mathbb{Z}_+^m$ , record its sparsity pattern; this point set is determined only by terms with non zero  $c_\alpha$ . The convex hull of this set of points is called the Newton Polytope of  $f$ , denoted  $NP(f) \subset \mathbb{R}^m$ . Consider now a system of  $m$  polynomials  $f_1 = \dots = f_m = 0$  in  $m$  variables and their Newton Polytopes,  $P_1, \dots, P_m$  ( $P_i = NP(f_i)$ ). The BKK theorem uses the mixed volume of these polytopes,  $MV(P_1, \dots, P_m)$ , to derive a bound on the number of non-zero complex solutions. The mixed volume is defined by the following formula

$$MV(P_1, \dots, P_m) = \sum_{k=1}^m (-1)^{m-k} \sum_{\substack{N \subseteq \{1, \dots, m\} \\ |N|=k}} \text{Volume} \left( \sum_{i \in N} P_i \right) \quad (10)$$

where we use  $\sum_{i \in N} P_i$  to denote the Minkowski sum of polytopes. In the special case in which all the polynomials share the same sparsity pattern, and hence the same Newton Polytope, i.e.,  $P_i = P$  for all  $i$ , then the mixed volume is given by  $MV(P_1, \dots, P_m) = m! \text{Volume}(P)$ . Indeed, (6) defines such a polynomial system. Computing the associated mixed volume (using the PHCpack solver [34]) yields a bound of 160 solutions. Calculation of the mixed volume (10) for our quadratic formulation (8)-(9) yields a tighter bound consisting of 64 complex solutions. (We note that [30]’s formulation – Eq. (11) in their paper – yields a non tight Bernstein’s bound of 80.)

We verified numerically that indeed all 64 complex solutions are attained. We used simulations to produce random camera matrices, as well as point matches, and used our method below to numerically solve the obtained polynomial systems (8)-(9). We obtain exactly 64 solutions in nearly all cases (due to numerical issues 0.1% of our trials produced 62-63 solutions). (We further verified this with homotopy continuation.) Naturally, not all of these solutions are real. The histogram plot in Figure 2 (left) shows the distribution of real solutions obtained over 1000 experiments. Our experiments yielded on average  $23.4 \pm 5.05$  real solutions. We are unfortunately unaware of a theoretical method to bound the number of real solutions below 64.

## 2.4. Solving the polynomial equations

Our next goal is to construct an efficient solution scheme to find all the real solutions of our polynomial equations, (8)-(9). A classical method for solving such systems is by using *resultants*. Such solutions are typically an order of magnitude more efficient than iterative methods such as homotopy continuation. Given a polynomial system of equations, its resultant is a polynomial in the coefficients of the system that vanishes if the equations share a common root.

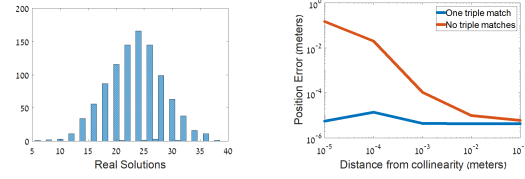


Figure 2: Left: A histogram plot showing the number of real solutions (of the 64 complex solutions). The plot represents the results of 1000 experiments in which camera settings and SIFT matches are randomly selected from the Herz-Jesus-P8 dataset. Right: Sensitivity analysis for three near collinear cameras. For a baseline of six meters between two cameras, the red curve shows the position error computed for the middle camera as a function of its deviation from the baseline. The blue line shows the position error when a single triple match is used to determine the unknown scale.

Resultants can further be used as an effective method for variable elimination. Treating one variable as a parameter (called a *hidden* variable), the resultant defines a polynomial equation in that variable, eliminating simultaneously all the rest of the variables. Various approaches can then be used to solve for the hidden variable and to extend the solution to all the rest of the variables.

The algebraic geometry literature offers several ways to construct resultants. Some of these approaches, unfortunately, may produce very large resultants that are difficult to work with. In this work we chose to use the Bezout-Cayley-Dixon (BCD) method [10], which allows us to obtain relatively compact expressions. We then use the obtained resultant to solve our equations by casting it as a generalized eigensystem problem in the form suggested in [26].

**The BCD method.** Given a polynomial system with  $m + 1$  equations in  $m + 1$  variables,  $x_1, \dots, x_{m+1}$ , and assume without loss of generality that we choose to hide  $x_{m+1}$ , then we express the polynomials in terms of the rest of the variables  $\mathbf{x} = (x_1, \dots, x_m)^T$  as

$$f_1(\mathbf{x}) = \dots = f_{m+1}(\mathbf{x}) = 0. \quad (11)$$

To construct the resultant we introduce new variables  $\mathbf{y} = (y_1, \dots, y_m)^T$  and construct the following  $(m+1) \times (m+1)$  matrix,  $D(\mathbf{x}, \mathbf{y}) =$

$$\begin{pmatrix} f_1(x_1, x_2, \dots, x_m) & f_2(x_1, x_2, \dots, x_m) & \dots & f_{m+1}(x_1, x_2, \dots, x_m) \\ f_1(y_1, x_2, \dots, x_m) & f_2(y_1, x_2, \dots, x_m) & \dots & f_{m+1}(y_1, x_2, \dots, x_m) \\ f_1(y_1, y_2, \dots, x_m) & f_2(y_1, y_2, \dots, x_m) & \dots & f_{m+1}(y_1, y_2, \dots, x_m) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(y_1, y_2, \dots, y_m) & f_2(y_1, y_2, \dots, y_m) & \dots & f_{m+1}(y_1, y_2, \dots, y_m) \end{pmatrix} \quad (12)$$

so that at each row  $i$ , one more variable  $y_i$  replaces a respective variable  $x_i$ . The Dixon polynomial is defined as

$$\delta(\mathbf{x}, \mathbf{y}) = \frac{\det(D(\mathbf{x}, \mathbf{y}))}{(x_1 - y_1) \cdots (x_m - y_m)}. \quad (13)$$

$\delta(\mathbf{x}, \mathbf{y})$  is indeed a polynomial; it can be verified that  $\det(D)$  is divisible by  $(x_1 - y_1) \cdots (x_m - y_m)$  by

the following argument. If we subtract row  $i + 1$  in  $D$  from row  $i$  ( $1 \leq i \leq m$ ; such an operation does not change the determinant) we obtain expressions of the form  $f_i(\dots, x_i, \dots) - f_i(\dots, y_i, \dots)$ . Such an expression vanishes at  $y_i = x_i$ , and so it contains a multiple of  $(x_i - y_i)$ . Overall,  $\delta(\mathbf{x}, \mathbf{y})$  is a polynomial of degree  $((i \times d_i) - 1)$  in  $x_i$  and  $((m + 1 - i) \times d_i - 1)$  in  $y_i$  for  $1 \leq i \leq m$ , where  $d_i$  is the maximal degree of  $x_i$  in  $f_1, \dots, f_{m+1}$ .

The Dixon polynomial,  $\delta(\mathbf{x}, \mathbf{y})$ , vanishes for any common root  $\mathbf{x}$  of (11), regardless of  $\mathbf{y}$ . This is clear, because the first row of  $D$ , which is independent of  $\mathbf{y}$ , vanishes with any common root  $\mathbf{x}$ . If we now express  $\delta(\mathbf{x}, \mathbf{y})$  as a sum of monomials in  $\mathbf{y}$ , i.e.,

$$\delta(\mathbf{x}, \mathbf{y}) = \sum_{\alpha} \left( \sum_{\beta} \tilde{M}_{\alpha, \beta} \mathbf{x}^{\beta} \right) \mathbf{y}^{\alpha} \quad (14)$$

then clearly all the coefficients of  $\mathbf{y}^{\alpha}$ ,  $\sum_{\beta} \tilde{M}_{\alpha, \beta} \mathbf{x}^{\beta}$ , must vanish identically. This yields the following linear system

$$\tilde{M} \tilde{\mathbf{v}} = 0, \quad (15)$$

where  $\tilde{M}$ , called the Dixon matrix, is the matrix of the coefficients  $\tilde{M}_{\alpha, \beta}$ , and  $\tilde{\mathbf{v}} = (\dots, \mathbf{x}^{\beta}, \dots)^T$ .

In principle, if  $\tilde{M}$  was square, we could solve for (11) by finding values of the hidden variable that make  $\tilde{M}$  singular (recall that its entries are polynomials in the hidden variable  $x_{m+1}$ ), and then computing the null space of  $\tilde{M}$ . In many cases, however,  $\tilde{M}$  is rectangular and may be identically rank deficient. Kapur et al. [18] showed that under certain conditions requiring the determinant of any maximal rank submatrix of  $\tilde{M}$  to vanish provides a necessary condition for the existence of a common root. This way, it is possible to obtain, after elimination of rows and columns, an informative square matrix  $M$ .

**Symbolic construction of the Dixon matrix.** Our polynomials (8)-(9) consist of 7 equations in 7 unknowns. We chose to hide  $q_2$  (recall that we set  $q_1 = 1$ ), yielding  $\mathbf{x} = (d_1, d_2, d_3, d_4, q_3, q_4)$ , and hence a matrix  $D$  of size  $7 \times 7$ . Numeric construction of the Dixon matrix can be obtained by plugging values for the 6 corresponding points into (8) and then applying the Maple package of [27] to  $D$ . Using the method above we obtain a Dixon matrix  $M$  of size  $27 \times 27$  whose entries are polynomials in  $q_2$ , with maximal degree 8, and a vector  $\mathbf{v}$  of size 27, composed of all powers of  $q_3 q_4$  of degree up to 6, except  $q_3^6$  (due to the order of elimination). Interestingly in this process all terms that involve  $\mathbf{d}$  are eliminated, and so our resultant only depends on the rotation parameters. This will be useful when the cameras are all collinear, as we further elaborate in Sec. 2.4.

To speed up computations we aim to generate a symbolic expression of the Dixon matrix  $M$ . Direct application of this construction unfortunately yielded polynomials that were too long to store in the computer memory. (Already

the determinant of a  $5 \times 5$  submatrix produced an expression of size 176Mb, and its calculation required even more memory.) To overcome this, we recursively applied the Laplace expansion, constructing polynomials from determinants of all submatrices of  $D$  of size  $3 \times 3$ . We then simplified these polynomials and introduced additional variables for the obtained coefficients of terms in  $\mathbf{v}$  and the hidden variable  $q_2$ . We later used the obtained expressions to construct polynomials for determinants of all submatrices of size  $4 \times 4$ . We repeated this process until we obtained a single polynomial expression for the determinant of the full  $7 \times 7$  matrix  $D(\mathbf{x}, \mathbf{y})$ . Next, we reorganized our terms in the form of (15), obtaining a matrix  $\tilde{M}$  of size  $27 \times 184$ . Finally, after removing all identically zero columns, we obtained a  $27 \times 27$  Dixon matrix  $M$  of the form

$$M(q_2) = M_0 + q_2 M_1 + q_2^2 M_2 + \dots + q_2^8 M_8, \quad (16)$$

where the entries of  $M_i$  ( $0 \leq i \leq 8$ ) are functions of known quantities,  $\mathbf{p}$ ,  $\mathbf{s}$ , and  $\mathbf{b}$ , (3), which are determined by each of the 6 corresponding pairs of points given as input. Following (15), this matrix satisfies

$$M \mathbf{v} = 0. \quad (17)$$

**Finding the roots.** To find the common roots of (11) we need to find assignments of the hidden variable,  $q_2$ , that make  $M$  singular. This can be achieved for example by calculating the roots of  $\det(M)$  as a univariate polynomial of degree 64. A more stable approach is to apply an eigen-decomposition technique, following [26]. Given (16), it can be readily verified that  $q_2$  is an eigenvalue of the following generalized eigen-system

$$C_2 \bar{\mathbf{v}} = q_2 C_1 \bar{\mathbf{v}}, \quad (18)$$

with

$$C_1 = \begin{pmatrix} I & 0 & \dots & 0 \\ 0 & I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & M_8 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 0 & I & 0 & \dots & 0 \\ 0 & 0 & I & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & I \\ -M_0 & -M_1 & -M_2 & \dots & -M_7 \end{pmatrix}$$

and the corresponding eigenvector  $\bar{\mathbf{v}}$  is given by

$$\bar{\mathbf{v}} = (\mathbf{v}; q_2 \mathbf{v}; q_2^2 \mathbf{v}; \dots; q_2^7 \mathbf{v}). \quad (19)$$

The matrices  $C_1$  and  $C_2$  are of size  $(27 \times 8) \times (27 \times 8) = 216 \times 216$ . Using our symbolic calculation we verified that  $\det(M) = \det(C_2 - q_2 C_1)$  is a polynomial of degree 64 and that  $C_1$  is singular. Consequently, of the 216 eigenvalues of this generalized eigen-system all but 64 of them diverge. This is compatible with the BKK bound derived in Sec. 2.3. Note that, since we have set  $q_1 = 1$ , a smaller number of solutions may be obtained if rotation of  $\pi$  is a solution. The remaining solutions can be found by repeating the process after rotating the global coordinate system.

Of the 64 eigenvalues we ignore complex ones. We then assign each of the remaining eigenvalues to  $q_2$  and extract  $q_3$  and  $q_4$  from the second and third entries of the respective eigenvector  $\bar{\mathbf{v}}$ . Together we obtain the quaternion  $\hat{\mathbf{q}} = (1, q_2, q_3, q_4)$ , which we then normalize to obtain a rotation quaternion. Note that there is no sign ambiguity since two unit quaternions  $\hat{\mathbf{q}}$  and  $-\hat{\mathbf{q}}$  represent the same rotation.

Once we recover the rotation parameters we use (6), which is linear in  $\mathbf{t}$ , to solve directly for  $\mathbf{t}$ . In a general camera setup this linear system is of rank 3, allowing us to solve for all the three degrees of freedom in  $\mathbf{t}$ . If this linear system is degenerate, however, it indicates that the cameras are all (near) collinear, in which case we only recover  $\mathbf{t}$  in a line in 3D (i.e., up to scale if we set the origin of the global coordinate system at any point along the line).

We can resolve this ambiguity in the translation by using one matching triplet, if available. Let  $\mathbf{p}_i \in I_i$ ,  $\mathbf{p}_j \in I_j$  ( $1 \leq i < j < n$ ), and  $\mathbf{p} \in I_n$ , and let  $\mathbf{P}$  denote the 3D point obtained by triangulating  $\mathbf{p}_i$  and  $\mathbf{p}_j$ . Suppose that  $\mathbf{t}$  is recovered in a line, i.e.,  $\mathbf{t} = \mathbf{t}' + \alpha \mathbf{t}''$ , where both  $\mathbf{t}', \mathbf{t}'' \in \mathbb{R}^3$  are known and  $\alpha$  is an unknown scalar. Then,

$$\mathbf{p} \propto R^T(\mathbf{P} - \mathbf{t}) = R^T(\mathbf{P} - \mathbf{t}' - \alpha \mathbf{t}''). \quad (20)$$

We can determine  $\alpha$  by solving the linear system

$$\mathbf{p} \times R^T(\mathbf{P} - \mathbf{t}' - \alpha \mathbf{t}'') = 0. \quad (21)$$

Note finally that with one triple match we need only 4 more pairs of points to solve for all six degrees of freedom in the exterior parameters. (See also [17] for solutions that combine pairs and triplets.) The triple match  $\mathbf{p}, \mathbf{p}_i, \mathbf{p}_j$  provides two polynomial equations to our polynomial system (3) (for the pairs  $\mathbf{p}, \mathbf{p}_i$  and  $\mathbf{p}, \mathbf{p}_j$ ), allowing with the additional 4 pairs to solve for 5 of the DOFs in the camera matrix. The same triple match can be used further to solve for the missing scale using (21).

Figure 2 (right) shows the error in recovering camera position for near collinear camera setups. Indeed, in this case, with just pairwise correspondences the position of the recovered camera is determined only up to the scalar  $\alpha$ . However, using a single matching triplet this scalar is recovered accurately even for exactly collinear cameras. Our method is summarized in Alg.1.

**More than 3 correspondences from one camera.** The  $27 \times 27$  Dixon matrix  $M$  constructed in (16) becomes singular when four of the six matching pairs come from a single image  $I_i$  ( $1 \leq i \leq n-1$ ) (and the other two from either one or two images), in which case its rank for a general assignment of  $q_2$  is 23. Following Kapur et al.'s main theorem [18], we use the sub-matrix  $\bar{M} = M_{1:23,1:23}$  consisting of the first 23 rows and columns of  $M$  which is generally full rank. The solutions for  $q_2$  can be found by applying the respective generalized eigen-system for  $\bar{M}$ .

To solve for the remaining variables  $q_3$  and  $q_4$ , we note first that, unlike in the previous case, the obtained eigenvector  $\bar{\mathbf{v}}$ , which satisfies  $\bar{M}\bar{\mathbf{v}} = 0$ , cannot be used directly toward this goal. Instead, we plug in each of the solutions to  $q_2$  into  $M(q_2)$ , obtaining a matrix of rank 22, and reorder its columns so that the columns that correspond to the monomials  $[1, q_4, q_4^2, q_4^3, q_4^4, q_4^5]^T$  of  $\mathbf{v}$  are placed on the right of  $M$ . We next apply the LU decomposition to  $M$ , so that  $M(q_2)\mathbf{v} = LU\mathbf{v} = 0$  which implies that  $U\mathbf{v} = 0$ . All but the 5 right most entries of the  $22^{th}$  row of  $U$  are zeros. Multiplying this row by  $\mathbf{v}$  (permuted accordingly) results in a 5-degree polynomial in the single variable  $q_4$ . For each of the 5 obtained solutions for  $q_4$  we once more reorder the columns of  $M(q_2)$  this time placing the columns corresponding to the monomials  $[1, q_4, q_4^2, q_4^3, q_4^4, q_3]^T$  of  $\mathbf{v}$  on the right. We next use the LU decomposition to obtain a linear equation in  $q_3$ . We finally plug in the 5 solutions for  $q_3$  and  $q_4$  into our original system of equations and discard all but the solution of minimum residual. (In practice for stability, we keep the solution that minimizes the Sampson error for the 6 corresponding points.) This case was solved in [37] and demonstrated empirically 40 solutions. We note, however, that [37]'s formulation (Eqs. (3-5) in their paper) yields a non tight BKK bound, which is 56. In our formulation, we achieve a tight BKK bound, which is 40, by placing the origin at the center of the camera which has 4 correspondences with the new camera.

We note finally that, as is mentioned in [37], when 5 correspondences come from the same camera we can first recover the essential matrix relating this camera to the new camera. This yields 20 solutions, which we can use to recover 5 of the 6 degrees of freedom in the pose of new camera. We then use the remaining matching pair to recover the remaining parameter (scale of translation).

---

**Algorithm 1:** Recover camera position and orientation for an image  $I_n$  given the locations and orientations of  $I_1, \dots, I_{n-1}$

---

**Input** :  $R_1, \dots, R_{n-1}, \mathbf{t}_1, \dots, \mathbf{t}_{n-1}, \{(\mathbf{p}_k, \mathbf{p}'_k)\}_{k=1}^6$   
*s.t.*  $\mathbf{p}_k \in I_n, \mathbf{p}'_k \in I_1 \cup \dots \cup I_{n-1}$

**Output:**  $\{U_i\}, \{\tau_i\}$ , a solution set of orientations and locations

- 1  $M_0, \dots, M_8 \leftarrow$  Use our symbolic derivations to construct resultant matrices (16)
  - 2  $C_1, C_2 \leftarrow$  Build generalized eigen-system (18)
  - 3  $\{(q_2, q_3, q_4)\}_{\{1..64\}} \leftarrow$  Solve eigen-system (18)
  - 4  $\{\hat{\mathbf{q}}_i\} \leftarrow$  Normalize and keep real solutions
  - 5  $\{U_i\} \leftarrow$  extract orientation matrices
  - 6  $\{\tau_i\} \leftarrow$  Solve for positions linearly using (6)
  - 7 return  $\{U_i\}, \{\tau_i\}$
-

Table 1: Runtime, orientation recovery error and residual obtained with our method compared to existing methods.

		General Case				4+2	
		Ours	Byrod et al.	Larsson et al.	Stewenius et al.	Ours	Zheng et al.
Runtime (ms)		21.4	260.1	<b>1.7</b>	2.3	20.3	<b>1.4</b>
Rotation Error (degrees)	mean	<b>6.3096e-07</b>	9.7899e-04	0.2457	1.8051	<b>0.0041</b>	4.2196
	median	<b>7.6592e-09</b>	5.9371e-08	3.3821e-05	0.1330	<b>7.1092e-08</b>	3.2804e-04
Stewenius' residual	mean	<b>4.3453e-10</b>	7.1570e-08	9.7387e-04	0.0045	<b>5.3624e-07</b>	9.9476e+09
	median	<b>3.5269e-12</b>	4.2803e-11	1.8555e-08	5.5276e-05	<b>6.5168e-11</b>	1.3399e-07

Table 2: Accuracy of pose recovery from landmarks. Results shown are averages over 22 different camera locations.

			3+3				4+2	
			Ours	Byrod et al.	Larsson et al.	Stewenius et al.	Ours	Zheng et al.
Position error	Meters	mean	<b>0.0278</b>	0.0323	0.5418	0.9837	<b>0.0148</b>	0.1164
		median	<b>0.0286</b>	<b>0.0286</b>	0.0388	0.1022	<b>0.0132</b>	0.0135
Orientation error	Frobenius	mean	<b>0.0063</b>	<b>0.0063</b>	0.1657	0.2213	<b>0.0050</b>	0.1097
		median	<b>0.0064</b>	<b>0.0064</b>	0.0074	0.0223	<b>0.0048</b>	0.0050
	Degrees	mean	<b>0.2538</b>	<b>0.2538</b>	6.8315	9.0557	<b>0.2020</b>	5.165
		median	<b>0.2599</b>	<b>0.2599</b>	0.3002	0.9037	<b>0.1938</b>	0.2010

### 3. Experiments

#### 3.1. Runtime and accuracy

We tested our solution in simulations and on real data. To assess its accuracy we used the Herz-Jesus-P8 dataset [31] to generate 1000 different configurations of six pairs of points (true matches) that include 1-3 matches from up to 6 cameras. We further generated 1000 6-pair configurations that involve 4 matches from one known camera. We compare our results for the former collection with results obtained with [30, 6, 21] and on the latter with [37]. (Recall that the solutions in [30, 6, 21] are degenerate when 4 matches come from one camera, while [37] only addressed that special case.)

The results are presented in Figure 3 and are summarized in Table 1. In each case we show the recovery error of camera orientation along with the residual error obtained when each solution (including ours) is plugged into the system of 30 equations of Stewenius et al. [30] in  $q_2$ ,  $q_3$  and  $q_4$ . (We do not show the errors in the location parameters since in all methods those are solved linearly once the orientation parameters are recovered.) Our method achieves highly accurate results outperforming these existing methods.

Our implementation utilizes C code to construct our resultant. The code obtains six candidate matches and their respective camera matrices, as described in Sec. 2.1, and uses our symbolic expressions (Sec. 2.4) to produce the nine  $27 \times 27$  matrices,  $M_0, \dots, M_8$  (16). We compiled the code as a Matlab Mex library and used MATLAB to solve the generalized eigen-system described in Sec. 2.4. Our solution runs in 20.3-21.4ms on a PC with i7-6700 3.4GHz CPU, which is compatible with real-time applications. Extracting Dixon matrix takes only 6ms of the 20ms, much faster than the general implementation [27] which takes about 17 seconds. Solving the generalized eigen-system of the sparse matrices  $C_1$  and  $C_2$  of size  $216 \times 216$  (18) takes 14ms of the 20ms. Our method is not as fast as the less accurate

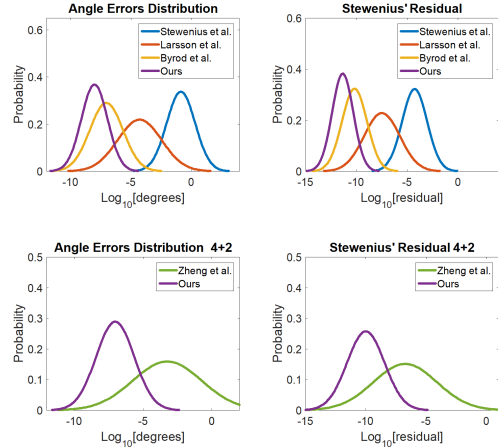


Figure 3: Distributions of orientation error and residuals produced (log10 scale) with our method compared to existing methods in 1000 trials.

methods in [30, 21, 37], but is significantly faster (and still more accurate) than [6] which is incompatible with real-time applications, see Table 1. For comparison, solving our equations with homotopy continuation takes 271ms using the multicore version of the PHCpack solver [34]. All these methods were run on the same PC.

#### 3.2. Landmark tracking

We simulated a landmark based robot positioning application. Our setup is described in Sec. 1.1 (see Fig. 1). Our goal is to recover the robot’s position and orientation as it moves, where at each time step we only use the robot’s current image and the two stationary images. To evaluate the methods we produced “ground truth” measurements by taking 78 images of the scene from multiple locations using a single camera with fixed internal parameters. We calibrated the images using VisualSFM [36], utilizing their EXIF tags,



Table 3: Accuracy of exterior parameter estimation using our sequential SFM pipeline compared to visualSFM. The table shows the mean and median over the cameras’ parameters error, where the error for each camera is averaged over 10 different runs.

			Fountain-P11		Herz-Jesus-P8		Castle-P30	
			Ours	VisualSFM	Ours	VisualSFM	Ours	VisualSFM
Position error	Meters	mean	<b>0.0025</b>	0.0027	<b>0.0042</b>	0.0046	<b>0.0222</b>	0.0316
		median	<b>0.0026</b>	0.0030	<b>0.0038</b>	0.0043	<b>0.0210</b>	0.0248
Orientation error	Frobenius	mean	<b>0.00065</b>	0.00067	<b>0.00042</b>	0.00054	<b>0.00098</b>	0.0014
		median	<b>0.00056</b>	0.00059	<b>0.00041</b>	0.00051	<b>0.00099</b>	0.0012
	Degrees	mean	<b>0.0263</b>	0.0271	<b>0.0172</b>	0.0219	<b>0.0399</b>	0.0555
		median	<b>0.0228</b>	0.0241	<b>0.0166</b>	0.0208	<b>0.0405</b>	0.0478

and constraining VisualSFM to share the same internal parameters across all the images. We further used the PTlens [1] software to remove radial distortion from the images. The overall scale was adjusted to meters by triangulating points with known distance that could be seen from some of the images. The obtained internal calibration was used for all the cameras in the experiment. The exterior calibration parameters were used to determine the positions and orientations of the two stationary cameras in the experiment, and as ground truth measurements to evaluate our estimated positions and orientations of the robot for the 22 tested images.

Table 2 shows the pose recovery errors obtained with our method both with 3 landmark points taken from each stationary camera and with 4 landmarks used from one stationary camera and 2 landmarks from the other camera. As the table shows, our method outperforms [30, 37, 21] and achieves comparable accuracies as [6], which is significantly slower than our method.

### 3.3. RANSAC iterations

The next experiment demonstrates that our improved accuracy can affect the number of needed RANSAC iterations. For this experiment we applied RANSAC to SIFT matches extracted from three images from the Fountain-P11 dataset. We assume we know the parameters of two of the three cameras and use RANSAC with our method to compute the orientation of the third camera. Each RANSAC iteration selects three matching pairs from each of the known camera. Our method than produces 64 complex solutions from which we select the solution that minimizes the sum of Sampson errors for the six matching pairs. We finally plot for each RANSAC iteration the minimal error with respect to ground truth obtained up to that iteration. Fig. 4 shows the accuracy achieved with this procedure, as a function of RANSAC iteration, compared with the accuracies achieved when our method is replaced by [30, 21]. It can be seen that our method achieved better accuracies in fewer iterations.

### 3.4. Sequential structure from motion

We finally demonstrates the utility of our method in a sequential multiview SFM pipeline. We produced a pipeline, similar to [29], but with RANSAC implemented with our method. We begin with two images and use [20]’s implementation of the 5-points algorithm to compute the essen-

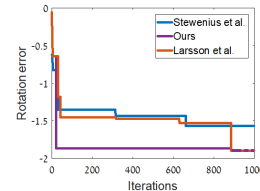


Figure 4: Camera orientation recovery using RANSAC. The Graph shows rotation error ( $\log_{10}(\text{degree})$ ) as a function of the number of RANSAC samples.

tial matrix between them. We then apply bundle adjustment [33] (using the SBA package [25]) to obtain the two corresponding camera matrices. Next, for each additional image we use our method, with RANSAC, utilizing the previously computed cameras to robustly recover the exterior parameters of the new camera. This was followed by bundle adjustment, as in [29], first using the already triangulated 3D points (using [15]), then triangulating new inlier matches, and, finally, removing points with large reprojection error.

We tested our pipeline on the Herz-Jesus-P8, fountain-P11, and castle-P30 datasets [31]. We compare our pipeline to VisualSFM [36]. For fair comparison we let both our method and VisualSFM use the same set of candidate matches (computed using [35]) and the same order of images in both pipelines. Results are shown in Table 3. Despite the use of matching pairs, instead of triplets, our method outperformed VisualSFM on all three datasets.

## 4. Conclusion

We have introduced a novel solution to the 6 point, online problem of camera pose estimation from pairwise matches by constructing the Dixon resultant, which we implemented efficiently through symbolic derivations. We further provided a rigorous proof showing that there are 64 roots for the underlying polynomial system. We showed in experiments that our method obtains more accurate solutions than existing methods and showed the utility of the method in sequential SFM pipeline. We plan in future work to develop an analogous method for uncalibrated images.

**Acknowledgment** Supported by the Minerva foundation with funding from the Federal German Ministry for Education and Research.



## References

- [1] PTLens automatic distortion correction. <http://epaperpress.com/ptlens/>.
- [2] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.
- [3] M.-A. Ameller, B. Triggs, and L. Quan. Camera pose revisited—new linear algorithms. In *European Conf. on Computer Vision (ECCV)*, 2000.
- [4] M. Arie-Nachimson, S. Z. Kovalsky, I. Kemelmacher-Shlizerman, A. Singer, and R. Basri. Global motion estimation from point matches. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 81–88, 2012.
- [5] D. N. Bernshtein. The number of roots of a system of equations. *Functional Analysis and its Applications*, 9(3):183–185, 1975.
- [6] M. Byröd, K. Josephson, and K. Åström. Fast and stable polynomial equation solving and its application to computer vision. *Int. Journal of Computer Vision*, 84(3):237–256, 2009.
- [7] F. Camposco, T. Sattler, A. Cohen, A. Geiger, and M. Pollefeys. Toroidal constraints for two-point localization under high outlier ratios. In *Proc. CVPR*, volume 1, 2017.
- [8] F. Camposco, T. Sattler, and M. Pollefeys. Minimal solvers for generalized pose and scale estimation from two rays and one point. In *European Conf. on Computer Vision*, pages 202–218. Springer, 2016.
- [9] D. A. Cox, J. Little, and D. O’shea. *Using algebraic geometry*, volume 185. Springer, 2006.
- [10] A. L. Dixon. The eliminant of three quantics in two independent variables. *Proc. of the London Mathematical Society*, 2(1):49–69, 1909.
- [11] I. Z. Emiris. *Sparse elimination and applications in kinematics*. PhD thesis, University of California, Berkeley, 1994.
- [12] X.-S. Gao, X.-R. Hou, J. Tang, and H.-F. Cheng. Complete solution classification for the perspective-three-point problem. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(8):930–943, 2003.
- [13] W. R. Hamilton. On quaternions; or on a new system of imaginaries in algebra. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 25(163):10–13, 1844.
- [14] B. K. Horn. Relative orientation revisited. *JOSA A*, 8(10):1630–1638, 1991.
- [15] Itseez. Open source computer vision library. <https://github.com/itseez/opencv>, 2015.
- [16] N. Jiang, Z. Cui, and P. Tan. A global linear method for camera pose registration. In *Proc. of the IEEE Int. Conf. on Computer Vision (ICCV)*, pages 481–488, 2013.
- [17] K. Josephson, M. Byrod, F. Kahl, and K. Astrom. Image-based localization using hybrid feature correspondences. In *Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [18] D. Kapur, T. Saxena, and L. Yang. Algebraic and geometric reasoning using dixon resultants. In *Proc. of the Int. Symposium on Symbolic and Algebraic Computation*, pages 99–107, 1994.
- [19] Z. Kukelova, M. Bujnak, and T. Pajdla. Automatic generator of minimal problem solvers. In *European Conf. on Computer Vision*, pages 302–315, 2008.
- [20] Z. Kukelova, M. Bujnak, and T. Pajdla. Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *British Machine Vision Conf.*, volume 2, page 2008, 2008.
- [21] V. Larsson, K. Åström, and M. Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [22] V. Larsson, M. Oskarsson, K. Åström, A. Wallis, Z. Kukelova, and T. Pajdla. Beyond gröbner bases: Basis selection for minimal solvers. *arXiv:1803.04360*, 2018.
- [23] V. Lepetit, F. Moreno-Noguer, and P. Fua. EPnP: An accurate O(n) solution to the PnP problem. *Int. Journal of Computer Vision*, 81(2):155–166, 2009.
- [24] H. Li and R. Hartley. Five-point motion estimation made easy. In *Int. Conf. on Pattern Recognition (ICPR)*, volume 1, pages 630–633, 2006.
- [25] M. I. Lourakis and A. A. Argyros. SBA: A software package for generic sparse bundle adjustment. *ACM Trans. on Mathematical Software (TOMS)*, 36(1):2, 2009.
- [26] D. Manocha. Solving systems of polynomial equations. *IEEE Computer Graphics and Applications*, 14(2):46–55, 1994.
- [27] M. Minimair. DR: Dixon Resultant Package for Maple. <http://minimair.org/dr>, 2015.
- [28] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 26(6):756–770, 2004.
- [29] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from internet photo collections. *Int. Journal of Computer Vision*, 80(2):189–210, 2008.
- [30] H. Stewenius, D. Nistér, M. Oskarsson, and K. Åström. Solutions to minimal generalized relative pose problems. In *OMNIVIS 2005*, 2005.
- [31] C. Strecha, W. Von Hansen, L. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.
- [32] C. Sweeney, V. Fragoso, T. Höllerer, and M. Turk. gdlS: A scalable solution to the generalized pose and scale problem. In *European Conf. on Computer Vision*, pages 16–31. Springer, 2014.
- [33] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment: a modern synthesis. In *Int. Workshop on Vision Algorithms*, pages 298–372, 1999.
- [34] J. Verschelde. Algorithm 795: Phcpack: A general-purpose solver for polynomial systems by homotopy continuation. *ACM Trans. on Mathematical Software (TOMS)*, 25(2):251–276, 1999.
- [35] C. Wu. Siftgpu: A gpu implementation of scale invariant feature transform (sift)(2007). URL <http://cs.unc.edu/~ccwu/siftgpu>, 2011.
- [36] C. Wu. Towards linear-time incremental structure from motion. In *3DTV*, pages 127–134, 2013.
- [37] E. Zheng and C. Wu. Structure from motion using structureless resection. In *Int. Conf. on Computer Vision (ICCV)*, pages 2075–2083, 2015.