

Partially calibrated semi-generalized pose from hybrid point correspondences

Snehal Bhayani¹ Torsten Sattler² Viktor Larsson³
Janne Heikkilä¹ Zuzana Kukelova⁴

¹Center for Machine Vision and Signal Analysis, University of Oulu, Finland

²Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague

³Computer Vision and Geometry Group, Department of Computer Science, ETH Zürich

⁴Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague

Abstract

In this paper we study the problem of estimating the semi-generalized pose of a partially calibrated camera, i.e., the pose of a perspective camera with unknown focal length w.r.t. a generalized camera, from a hybrid set of 2D-2D and 2D-3D point correspondences. We study all possible camera configurations within the generalized camera system. To derive practical solvers to previously unsolved challenging configurations, we test different parameterizations as well as different solving strategies based on state-of-the-art methods for generating efficient polynomial solvers. We evaluate the three most promising solvers, i.e., the **H51f** solver with five 2D-2D correspondences and one 2D-3D correspondence viewed by the same camera inside generalized camera, the **H32f** solver with three 2D-2D and two 2D-3D correspondences, and the **H13f** solver with one 2D-2D and three 2D-3D correspondences, on synthetic and real data. We show that in the presence of noise in the 3D points these solvers provide better estimates than the corresponding absolute pose solvers.

1. Introduction

Estimating camera geometry, i.e., absolute or relative pose and internal camera calibration, is a fundamental problem in computer vision with many applications, e.g., in camera calibration [35, 47], structure-from-motion (SfM) [34, 36, 41, 45], localization [3, 31, 32], visual odometry [25, 26], and image retrieval [29, 44].

Camera geometry solvers are usually used inside RANSAC-style hypothesis-and-test frameworks [10]. For efficiency, it is therefore important to employ minimal solvers that generate the solution from a minimal number of point correspondences. Minimal relative and absolute pose problems have been extensively studied for decades with many solutions for calibrated cameras [10, 27, 28],

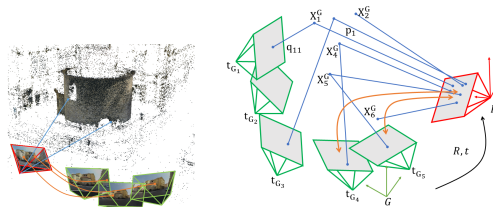


Figure 1. An illustration of the problem configuration.

partially calibrated cameras with unknown focal length [4, 19, 20, 22, 38], cameras with unknown radial distortion [6, 17, 18, 22], and solutions assuming known gravity direction [9, 16, 33]. The minimal solvers to these problems are based on different parameterizations and different solution methods. E.g., the absolute pose problem for a camera with unknown focal length has been solved based on ratio of distances [4], the 3.5pt formulations from [22, 46], and a solution based on the Cayley parameterization of rotation that is solved using the extremely efficient 3Q3 solver [19].

Generalized camera: All above-mentioned algorithms assume the central perspective projection model (potentially with radial distortion). Recently, several minimal solutions for generalized cameras were proposed. A generalized camera [30] can be represented by a arbitrary set of rays with, in general, different projection centers. This camera model has many applications, e.g., in SfM [48] and visual localization [37, 43], where we either work with multi-camera systems or we have to register a new camera (a query image) to a set of cameras with known poses modeled as a generalized camera composed of the known perspective ones. The latter case is called the semi-generalized pose estimation problem. In many situations, estimating the camera pose w.r.t. the generalized camera leads to more accurate poses compared to estimates from pair-wise epipolar geometries, especially thanks to a larger field-of-view of the generalized camera [40]. Moreover, in contrast to pinhole epipolar geometry, the scale of the translation can be recovered.

While the problem of estimating the absolute pose of a generalized camera can be solved very efficiently using the 3Q3 solver [19] (the final solver runs in a few μs), the problem of estimating the relative pose of two generalized cameras is significantly more complex [39]. This problem results in a complex system of 15 polynomial equations with 64 solutions and with a Gröbner basis solver [39] that is infeasible for real-time applications. Recently, several solutions to different semi-generalized relative pose problems were proposed. In [48], the authors considered a semi-generalized epipolar geometry problem, *i.e.*, the problem of estimating the relative pose together with the scale of the translation between one perspective and one generalized camera from 2D-2D correspondences. Due to the complex geometry of the problem, minimal solutions to only four different configurations were presented, *i.e.*, the E_{5+1} and E_{4+2} solvers for calibrated pinhole cameras, and the $E_{f_{6+1}}$ and $E_{f_{5+2}}$ solvers for pinhole cameras with unknown focal length. Here, a generalized camera consists of multiple perspective ones. 4+2 denotes a configuration where four point correspondences come from one of these perspective cameras and the remaining two correspondences come from one or two other perspective cameras. The proposed E_{4+2} and $E_{f_{5+2}}$ solvers were impractical for real-time applications since they perform eliminations of huge matrices and have running times of 1.2ms and 13.6ms, respectively. Recently, [2] showed that for planar scenes, the semi-generalized relative pose problem can be, after using suitable parameterizations and after eliminating some unknowns, solved efficiently by finding the roots of a single variable polynomial. The authors proposed such efficient minimal solutions for calibrated and partially-calibrated semi-generalized homography estimation and all different configurations of 2D-2D correspondences in a generalized camera.

Hybrid correspondences: In [13], the authors suggested to use combinations of 2D-2D and 2D-3D correspondences, *i.e.*, hybrid correspondences, for visual localization. Using hybrid correspondences has several advantages. While a 2D-2D correspondence provides only one constraint on the camera geometry, a 2D-3D correspondence provides two constraints and therefore decreases the number of correspondences needed for pose estimation. On the other hand, 3D points may not be available for many 2D detections in the query image, *i.e.*, it is not possible to triangulate these points, or these 3D points can be noisy. Thus, a combination of 2D-2D and 2D-3D correspondences may bring a benefit from both and, as shown in [7], can result in better pose estimates. [13] listed all possible minimal configurations of hybrid point correspondences for semi-generalized pose estimation for calibrated, partially calibrated (unknown focal length) and uncalibrated perspective cameras w.r.t. a generalized camera. While the authors estimated the number of solutions for all cases, they proposed solvers for only sev-

eral simple configurations, including configurations where all 2D points are observed by one camera inside the generalized camera or the calibrated case with two 2D-2D and two 2D-3D correspondences. They also suggested a solution to the case of uncalibrated cameras with one 2D-2D correspondences and five 2D-3D correspondences.

Recently, [7] proposed minimal solutions to several problems of estimating the pose of calibrated cameras, and cameras with known vertical direction, from hybrid point correspondences. These solutions assume that both cameras are generalized, *i.e.*, a fully generalized case. The final solvers were mostly obtained using the automatic generator based on Gröbner bases [15]. Together with the proposed minimal solvers for calibrated cameras, a RANSAC-based approach that is automatically selecting the “best” type of solver for each RANSAC iteration was presented. The solver to be used in the next iteration is selected in a data-driven way using a probability-guided sampling strategy, allowing it to adapt to the quality of the provided correspondences. The paper showed that properly combining different types of correspondences and different camera pose solvers for such correspondences can bring a significant benefit in the performance of RANSAC. Even though some of the proposed solvers are efficient, even for calibrated cameras there already are configurations that result in large solvers, *e.g.*, the solver that uses four 2D-2D and one 2D-3D correspondence and which has to perform elimination of a matrix of size 244×277 .

In this paper, we study challenging unsolved problems for estimating the *semi-generalized pose of a partially calibrated camera*, *i.e.*, the pose of a perspective camera with unknown focal length w.r.t. a generalized camera, *from a hybrid set of 2D-2D and 2D-3D point correspondences*. The proposed solvers fill the gaps that still remain in the arsenal of minimal solvers and provide new alternatives for pose estimation of a camera with unknown focal length¹ that can be efficiently used inside the hybrid RANSAC framework [7]. We assume a semi-generalized case, compared to the fully generalized case considered in [7], since this scenario appears more often in applications, *e.g.*, in visual localization, and results in simpler and faster solvers. The main **contributions** of the paper are:

1. We propose solutions to all possible minimal point configurations for semi-generalized pose estimation of a partially calibrated camera from a hybrid set of 2D-2D and 2D-3D point correspondences. The proposed solvers include (i) the H_{51f} solvers with five 2D-2D correspondences and one 2D-3D correspondence, (ii) the H_{32f} solvers with three 2D-2D and two 2D-3D correspondences, and (iii) the H_{13f} solvers with one 2D-2D correspondence and three 2D-3D correspon-

¹Note that in many applications, the only intrinsic parameter of a fully uncalibrated camera that needs to be estimated is the unknown focal length.

dences. In all three cases we consider all possible camera configurations within the generalized camera.

2. To derive efficient and stable solvers for all challenging configurations, we test different parameterizations of the problem *e.g.*, using the essential matrix, homographies, quaternions, and the Cayley parameterization of rotation, as well as different solution strategies based on state-of-the-art methods for generating efficient polynomial solvers, *e.g.*, the elimination ideal method [20], heuristic-based basis sampling approach [23] and resultants [1]².
3. We test the most practical solvers on synthetic as well as real data. We show that in the presence of noise in 3D points and for special type of motions, *e.g.*, forward motion, these solvers provide better estimates than corresponding absolute pose solvers.

2. Problem formulation

Let us consider a camera setup as depicted in Fig. 1. We denote the pinhole query camera as \mathcal{P} and the generalized camera as \mathcal{G} . The generalized camera \mathcal{G} is assumed to be fully calibrated, and it consists of a set of pinhole cameras denoted as $\{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_k\}$. In this paper, we consider \mathcal{P} to be partially calibrated. Its calibration matrix is of the form $K = \text{diag}(f, f, 1)$ with unknown focal length f . We use the upper index to denote a coordinate system. We consider two different coordinate systems for the generalized camera \mathcal{G} : the local coordinate system of the generalized camera \mathcal{G} as a single entity, and the local coordinate systems of each of the pinhole cameras, \mathcal{G}_i . Let $R_{G_i}, \mathbf{t}_{G_i}$ denote the rotation and translation required to align the local coordinate system of \mathcal{G}_i to the local coordinate system of the generalized camera \mathcal{G} . Let R_G, \mathbf{t}_G denote the rotation and translation required to align the local coordinate system of \mathcal{G} to the local coordinate system of \mathcal{P} : Let $\mathbf{X}^P \in \mathbb{R}^3$ and $\mathbf{X}^G \in \mathbb{R}^3$ be the coordinates of the point \mathbf{X} in the local coordinate system of \mathcal{P} and the local coordinate system of \mathcal{G} , respectively. It holds that $\mathbf{X}^P = R_G \mathbf{X}^G + \mathbf{t}_G$.

For such a camera setup, our goal is to estimate the rotation $R_G \in \mathbf{SO}(3)$ and the translation $\mathbf{t}_G \in \mathbb{R}^3$ between the generalized camera \mathcal{G} and the perspective camera \mathcal{P} , *i.e.*, to align the local coordinate system of \mathcal{G} with the local coordinate system of \mathcal{P} . Additionally we also need to estimate the focal length f of the camera \mathcal{P} . For the sake of brevity, we replace R_G with R and \mathbf{t}_G with \mathbf{t} .

Let us assume a 3D point \mathbf{X}_j observed by the perspective camera \mathcal{P} and the camera \mathcal{G}_i , *i.e.*, the i -th constituent perspective camera from the generalized camera \mathcal{G} .

²Finding a feasible formulation, *i.e.*, one leading to a practical solution, and deriving an efficient and stable solver to the resulting polynomial system usually requires many non-trivial “tricks” and a good knowledge of both camera and algebraic geometry. Thus, studying different formulations and testing different solution strategies itself is an important contribution.

Let us denote the image points detected in \mathcal{P} and \mathcal{G}_i as $\mathbf{p}_j = [x_j, y_j, 1]^\top$ and $\mathbf{g}_{ij} = [x_j^{G_i}, y_j^{G_i}, 1]^\top$, respectively. With this notation, the coordinates of the 3D point \mathbf{X}_j in the local coordinate system of \mathcal{P} are

$$\mathbf{X}_j^P = \alpha_j K^{-1} \mathbf{p}_j, \quad (1)$$

where K is the calibration matrix of the camera \mathcal{P} and α_j represents the depth of the point \mathbf{X}_j in \mathcal{P} . A similar relationship holds for the coordinates of the 3D point \mathbf{X}_j in the local coordinate system of \mathcal{G}_i as

$$\mathbf{X}_j^{G_i} = \beta_{ij} K_{G_i}^{-1} \mathbf{g}_{ij}, \quad (2)$$

where K_{G_i} is the calibration matrix of the camera \mathcal{G}_i and β_{ij} represents the depth of the point \mathbf{X}_j in \mathcal{G}_i . To obtain the relationship between \mathbf{X}_j^P and $\mathbf{X}_j^{G_i}$ we have to transform them into the same coordinate system, *i.e.*, in this case the local coordinate system of \mathcal{P} . This gives us

$$R(\beta_{ij} R_{G_i} K_{G_i}^{-1} \mathbf{g}_{ij} + \mathbf{t}_{G_i}) + \mathbf{t} = \alpha_j K^{-1} \mathbf{p}_j. \quad (3)$$

Note that here we use the fact that $R = R_G$ and $\mathbf{t} = \mathbf{t}_G$. Since in our case $R_{G_i}, \mathbf{t}_{G_i}$ and K_{G_i} are known, for better readability we substitute $\mathbf{q}_{ij} = R_{G_i} K_{G_i}^{-1} \mathbf{g}_{ij}$ and obtain

$$R(\beta_{ij} \mathbf{q}_{ij} + \mathbf{t}_{G_i}) + \mathbf{t} = \alpha_j K^{-1} \mathbf{p}_j. \quad (4)$$

This denotes the constraint imposed by a 2D-2D correspondence $\mathbf{p}_j \leftrightarrow (\mathbf{q}_{ij}, \mathbf{t}_{G_i})$. Similarly, if we have a 2D-3D correspondence between a 2D point \mathbf{p}_j and a 3D point \mathbf{X}_j^G in the local coordinate system of the generalized camera \mathcal{G} , then the resulting constraint is

$$R \mathbf{X}_j^G + \mathbf{t} = \alpha_j K^{-1} \mathbf{p}_j. \quad (5)$$

3. Minimal solvers

The problem of semi-generalized pose from hybrid point correspondences has seven degrees of freedom (d.o.f.), three each for R and \mathbf{t} , and one for f . From the constraint (4) induced by each 2D-2D point correspondence, we can eliminate the depths α_j and β_{ij} to obtain

$$(\mathbf{p}_j)^\top [KR \mathbf{q}_{ij}]_\times (KR \mathbf{t}_{G_i} + K \mathbf{t}) = 0, \quad (6)$$

where the notation $[\mathbf{a}]_\times$ indicates the skew-symmetric matrix of the vector $\mathbf{a} \in \mathbb{R}^3$. Thus, each 2D-2D point correspondence gives us one equation. Similarly, from the constraint (5) induced by each 2D-3D point correspondence, we can eliminate the depth α_j to obtain

$$[\mathbf{p}_j]_\times (KR \mathbf{X}_j^G + K \mathbf{t}) = \mathbf{0}, \quad (7)$$

which gives us two linearly independent equations.

A hybrid point configuration (m, n) consists of m 2D-2D and n 2D-3D point correspondences. It results in a total

of $m+2n$ linearly independent equations. Since in this case we have 7 d.o.f., for any hybrid point configuration (m, n) to lead to a minimal problem, we require that $m + 2n = 7$. We denote the solver for such a hybrid point configuration (m, n) as $\mathbf{H}mnf$ in this paper.

A given hybrid point configuration (m, n) can have different configurations of the generalized camera \mathcal{G} , based on the largest number $k \leq m$ of 2D-2D correspondences detected by the same pinhole camera \mathcal{G}_i within the generalized camera \mathcal{G} . For brevity we denote such a case as $[k]$.

Parameterizations: For all configurations we generated solvers using three different types of parameterizations:

- *Rotation & translation (R&t):* This parameterization correspond to Eqs. (6) and (7). Here we tested Cayley and quaternion-based parameterizations of the rotation matrix R . For the quaternion representation, inspired by [49], we use a four variable reparameterization of the product KR . This is important for removing symmetries and halving the number of solutions. We denote the initial polynomial system as E .
- *Homography (H):* We tested this parameterization for all configurations with at least three 2D-2D correspondences coming from the same camera \mathcal{G}_i^3 or three 2D-3D correspondences. Such three correspondences define a plane and thus a 3×3 homography matrix H induced by this plane. Three correspondences give us six linear equations in the elements of H and thus can be used to parameterize H using a three dimensional null space. Here, the initial system E is defined by this parameterization, together with equations coming from the decomposition of the homography matrix H and the equations from the remaining correspondences.
- *Fundamental matrix (F):* We tested this parameterization for configurations where we have five or six 2D-2D correspondences coming from the same camera \mathcal{G}_i^4 . Since each such correspondence gives us a linear constraint on the fundamental matrix F , we can parameterize it using a four resp. three dimensional null space. Here, the initial system E is defined by this parameterization, together with the equations coming from the decomposition of F and those induced by the remaining point correspondences.

To simplify the initial system of equations E in each configuration and parameterization, we tried to eliminate different variables using the elimination ideal method [20]. We also tried out different transformations of the coordinate systems for the generalized camera \mathcal{G} and the pinhole camera \mathcal{P} . To solve the resulting system of polynomial equations,

³Note, that here we can project 3D points to the camera \mathcal{G}_i to obtain a 2D-2D point correspondence. This means that the only case where this parameterization is not applicable is $\mathbf{H51f}[1]$.

⁴We can again project 3D points to the camera \mathcal{G}_i to obtain a 2D-2D point correspondence.

| Problem | #sols | Gen. cam. | GB [23] | Res. [1] | Param. | 2D-2D | 2D-3D |
|-----------------|-------|-------------|------------------|------------------|--------|-------|-------|
| $\mathbf{H13f}$ | 12 | [1] | 70×82 | 91×91 | H | 1 | 3 |
| $\mathbf{H32f}$ | 26 | [1],[2],[3] | 289×315 | — | R&t | 3 | 2 |
| $\mathbf{H51f}$ | 56 | [1] | 506×562 | 537×537 | R&t | 5 | 1 |
| $\mathbf{H51f}$ | 50 | [2],[3] | 511×561 | 374×374 | R&t | 5 | 1 |
| $\mathbf{H51f}$ | 38 | [4] | 390×428 | 243×243 | R&t | 5 | 1 |
| $\mathbf{H51f}$ | 9 | [5] | 9×18 | 10×10 | F | 5 | 1 |

Table 1. Partially calibrated semi-generalized pose solvers for hybrid point correspondences. $\mathbf{H}mnf$ denotes a case with m 2D-2D point correspondences and n 2D-3D point correspondences. $[k]$ denotes a case where no more than k 2D-2D correspondences are detected by a camera \mathcal{G}_i within the generalized camera \mathcal{G} .

we used two state-of-the-art algebraic methods for generating efficient polynomial solvers, *i.e.*, the Gröbner basis method [21] including the basis sampling strategy [23] and the hidden variable resultant-based method [1].

This results in huge number of different combinations of different parameterizations, point configurations, solution strategies, and solvers that we generated. In the paper, we present the fastest solvers among all generated solvers for each hybrid point configuration (m, n) and all possible generalized camera configurations $[k]$. These solvers are summarized in Tab. 1. Note that we have not studied the $(7, 0)$ case because our goal in this paper is to study hybrid point configurations with at least one 2D-3D point correspondence. The first two columns in this table report the solver/problem name and the number of solutions⁵ of the given configuration and its particular formulation, the third column lists the feasible configurations of cameras in a generalized camera that can be solved using the particular solver, the fourth and the fifth columns list the size of the smallest/fastest solver, generated using the basis sampling strategy [23] (elimination template matrix size), resp. the resultant-based method [1] (Generalized Eigenvalue Problem size), the sixth column is the parameterization of the problem that leads to this solver, and the last two columns depict the hybrid point configuration.

Next, we describe the fastest solvers for each studied hybrid point configuration, *i.e.*, $\mathbf{H13f}$, $\mathbf{H32f}$, and $\mathbf{H51f}$.

3.1. $\mathbf{H13f}$

In this case we have one 2D-2D point correspondence $\mathbf{p}_1 \leftrightarrow (\mathbf{q}_{11}, \mathbf{t}_{G_1})$ and three 2D-3D point correspondences $\mathbf{p}_j \leftrightarrow \mathbf{X}_j^G$, $j = 2, \dots, 4$, and therefore only one hybrid point configuration, *i.e.*, the configuration [1]. For this configuration, the parameterization that led to the smallest solver is the *Homography parameterization*.

In this case, the three 3D points \mathbf{X}_j^G , $j = 2, \dots, 4$, define

⁵Note that for some problems, the reported number of solutions does not correspond to the number of solutions presented in [13]. The reason for this is two-fold, first of all [13] did not consider possible symmetries, and second, in some problems the authors computed solutions to the case where both the perspective and the generalized camera have a common unknown focal length. However, this scenario is impractical.

a plane in the local coordinate system of the generalized camera \mathcal{G} . Let us denote this plane as π and its vector as \mathbf{N} , encoding both the direction of the plane normal and the distance from the origin. Therefore

$$\mathbf{X}_j^{\mathcal{G}} \in \pi \implies \mathbf{N}^\top \mathbf{X}_j^{\mathcal{G}} + 1 = 0, \quad j = 2, \dots, 4. \quad (8)$$

Coordinate system transform: W.l.o.g., we can rotate and translate the coordinate system of \mathcal{G} such that $\mathbf{N} = [0 \ 0 \ d]^\top$, $d \neq 0$ and $\mathbf{t}_{\mathcal{G}_1} = [0 \ 0 \ 0]^\top$. W.l.o.g., we can also rotate the coordinate system of the pinhole camera \mathcal{P} such that $\mathbf{q}_{11} = [1 \ 0 \ 1]^\top$.

Null-space parameterization of \mathbf{H}_K : Let us define the homography induced by the three 2D-3D point correspondences between the coordinate systems of \mathcal{G} and \mathcal{P} via point transfer through the plane π . Denoting the homography matrix as $\mathbf{H} \in \mathbb{R}^{3 \times 3}$ we can write

$$\mathbf{H} = \mathbf{R} - \mathbf{t}\mathbf{N}^\top. \quad (9)$$

The constraints imposed by a 2D-3D point correspondence, of the form (5), become

$$\mathbf{H}\mathbf{X}_j^{\mathcal{G}} = \alpha_j \mathbf{K}^{-1} \mathbf{p}_j, \quad j = 2, \dots, 4. \quad (10)$$

Let us write $\mathbf{t}_K = \mathbf{K}\mathbf{t}$ and $\mathbf{H}_K = \mathbf{K}\mathbf{H}$. Then Eqs. (9) and (10) can be rewritten as

$$\mathbf{H}_K = \mathbf{K}(\mathbf{R} - \mathbf{t}\mathbf{N}^\top) = \mathbf{K}\mathbf{R} - \mathbf{t}_K\mathbf{N}^\top \quad (11)$$

$$\mathbf{H}_K \mathbf{X}_j^{\mathcal{G}} = \alpha_j \mathbf{p}_j, \quad j = 2, \dots, 4. \quad (12)$$

Eq. (12) gives us six linear equations in the nine elements of \mathbf{H}_K . From these equations we can obtain a three-dimensional nullspace parameterization of the matrix \mathbf{H}_K . Thus, we can express \mathbf{H}_K as a function of three unknown variables, n_1 , n_2 and n_3 . In fact, we can express \mathbf{H}_K as

$$\mathbf{H}_K = n_1 \mathbf{N}_1 + n_2 \mathbf{N}_2 + n_3 \mathbf{N}_3, \quad (13)$$

where \mathbf{N}_1 , \mathbf{N}_2 and \mathbf{N}_3 are the matrix forms of the basis vectors of the null space.

Constraint on \mathbf{H}_K : Now we derive constraints on the matrix \mathbf{H}_K . We can rewrite Eq. (11) as

$$\mathbf{R} = \mathbf{K}^{-1}(\mathbf{H}_K + \mathbf{t}_K\mathbf{N}^\top). \quad (14)$$

Since $\mathbf{R} \in \mathbf{SO}(3)$, we have $\mathbf{R}\mathbf{R}^\top = \mathbf{R}^\top\mathbf{R} = \mathbf{I}$. Thus we have the following set of constraints from Eq. (14):

$$\begin{aligned} \mathbf{K}^{-1}(\mathbf{H}_K + \mathbf{t}_K\mathbf{N}^\top)(\mathbf{H}_K + \mathbf{t}_K\mathbf{N}^\top)^\top \mathbf{K}^{-\top} &= \mathbf{I}, \\ (\mathbf{H}_K + \mathbf{t}_K\mathbf{N}^\top)^\top \mathbf{K}^{-\top} \mathbf{K}^{-1}(\mathbf{H}_K + \mathbf{t}_K\mathbf{N}^\top) &= \mathbf{I}. \end{aligned} \quad (15)$$

Due to the proposed coordinate system transform, the constraint (6) for the 2D-2D point correspondence becomes

$$\mathbf{p}_1^\top [\mathbf{K}\mathbf{R}\mathbf{q}_{11}]_{\times} \mathbf{K}\mathbf{t} = 0. \quad (16)$$

Substituting \mathbf{R} from Eq. (14) into Eq. (16), we have

$$((\mathbf{H}_K + \mathbf{t}_K\mathbf{N}^\top)\mathbf{q}_{11})^\top [\mathbf{t}_K]_{\times} \mathbf{p}_1 = 0. \quad (17)$$

The equations in (15) and (17) are the constraints on the unknown quantities, \mathbf{H}_K , \mathbf{t}_K , and \mathbf{K} , and the known quantities, $\mathbf{N} = [0 \ 0 \ d]^\top$ and $\mathbf{p}_1 = [x_1 \ y_1 \ 1]^\top$. Let us define the ideal generated by these equations as $I \subset \mathbb{C}[\varepsilon]$ [8], where ε contains the nine unknowns from \mathbf{H}_K , three from \mathbf{t}_K , the inverse of the focal length $w = \frac{1}{f}$, and d , x_1 , and y_1 . Note that x_1 , y_1 , and d are known and here we treat them as known symbolic variables. Now, we can use the elimination ideal technique [20] to eliminate three unknowns of \mathbf{t}_K and w from this ideal. *I.e.*, we compute an elimination ideal I_1 that contains polynomials only in nine unknown variables from \mathbf{H}_K and three known variables, d , x_1 , and y_1 . Note, that this elimination ideal can be computed offline using some algebraic geometry software like Macaulay 2 [11]. We found that such an elimination ideal is generated by eight polynomials (two of degree 3, one of degree 4, and five of degree 8) in 12 variables (nine unknown and three known). For more details on elimination ideals see [8, 20].

Substituting the three-variable parameterization of \mathbf{H}_K from Eq. (13) into the generators of the ideal I_1 , we obtain a system of eight equations in three unknowns. Using Macaulay 2 [11], we verified that this system has up to 12 solutions. This system defines the minimal formulation for the **H13f** case. We used two state-of-the-art algebraic methods, *i.e.*, the Gröbner basis-based automatic generator [21] including the basis sampling strategy [23] as well as the resultant-based generator [1], to generate solvers for this system of eight equations in three unknowns. We found that smaller solvers can be obtained if, instead of using all eight equations, we use only six equations (two of degree 3, one of degree 4, and three of degree 8). Using the Gröbner basis method with the basis sampling strategy [23], the generated solver was of size 77×89 , while the one generated using the resultant-based method [1] was of size 91×91 . We report these solver sizes in Tab. 1.

Extracting pose from \mathbf{H} : Once we have computed the solutions to \mathbf{H}_K , we can estimate the value of $w = 1/f$ from the Eq. (15) through variable elimination and substitution. From w , we can then compute the calibration matrix \mathbf{K} and subsequently the homography matrix $\mathbf{H} = \mathbf{K}^{-1}\mathbf{H}_K$. An important step here is to efficiently extract the relative pose, *i.e.*, the rotation matrix \mathbf{R} and the translation vector \mathbf{t} from \mathbf{H} . Our chosen coordinate system transformation plays a crucial role here. Substituting $\mathbf{N} = [0 \ 0 \ d]^\top$ in Eq. (9), we have

$$\mathbf{R} = \mathbf{H} + \mathbf{t} [0 \ 0 \ d]. \quad (18)$$

Writing $\mathbf{H} = [\mathbf{h}_1 \ \mathbf{h}_2 \ \mathbf{h}_3]$, the rotation matrix \mathbf{R} can be computed as

$$\mathbf{R} = [\mathbf{h}_1 \ \mathbf{h}_2 \ [\mathbf{h}_1]_{\times} \mathbf{h}_2]. \quad (19)$$

Again, from Eq. (18) and the computed values of \mathbf{H} , the translation vector \mathbf{t} can be computed as

$$\mathbf{t} = \frac{1}{d}([\mathbf{h}_1]_{\times} \mathbf{h}_2 - \mathbf{h}_3) . \quad (20)$$

3.2. H32f

In the second configuration, we have three 2D-2D point correspondences $\mathbf{p}_j \leftrightarrow (\mathbf{q}_{ij}, \mathbf{t}_{G_i})$ and two 2D-3D point correspondences $\mathbf{p}_j \leftrightarrow \mathbf{X}_j^G$. Thus, there are three possible generalized camera configurations, *i.e.*, [1], [2] and [3]. For all three camera configurations, we obtained the smallest solvers using the *Rotation & translation* parameterization. Such a parameterization for H32f was already proposed in [13], where the authors suggested to use a coordinate transformation resulting in a system of five equations in five variables with 52 solutions. This system is, however, complex and it leads to a huge solver⁶. In order to generate a feasible solver with fewer solutions, in this paper, we reparameterized the product KR. Next, we review the coordinate transformation and the parameterization proposed in [13] and then describe our proposed solver. We obtain the same solver for all three configurations [1], [2] and [3].

Coordinate system transform: W.l.o.g., we can translate and rotate the coordinate system of \mathcal{G} such that $\mathbf{X}_4^G = [0 \ 0 \ 1]^\top$ and $\mathbf{X}_5^G = [0 \ 0 \ 0]^\top$. Moreover, we can rotate the coordinate system of the pinhole camera \mathcal{P} such that $\mathbf{p}_5 = [1 \ 0 \ 1]^\top$.

By this coordinate system transform, the constraint (5) for the fifth point correspondence $\mathbf{p}_5 \leftrightarrow \mathbf{X}_5^G$ leads to

$$\mathbf{R} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + \mathbf{t} = \alpha_5 \mathbf{K}^{-1} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \implies \mathbf{Kt} = \alpha_5 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} . \quad (21)$$

Substituting the above expression of \mathbf{Kt} in the constraints for the 2D-2D point correspondences and the fourth 2D-3D point correspondence, we obtain

$$(\mathbf{KRp}_j)^\top [\mathbf{q}_{ij}]_{\times} (\mathbf{KRt}_{g_i} + \alpha_5 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}) = \mathbf{0} , \quad (22)$$

$$[\mathbf{q}_4]_{\times} (\mathbf{KR}\mathbf{X}_4 + \alpha_5 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}) = \mathbf{0} , \quad (23)$$

where $i \in \{1, 2, 3\}$ and $j = 1, \dots, 3$. The constraints (22) and (23) together give us a set of five equations in the quantities \mathbf{K} , \mathbf{R} , and α_5 . We parameterized the rotation matrix \mathbf{R} using quaternions. We have five unknowns, one for the calibration matrix \mathbf{K} , three for the rotation matrix \mathbf{R} and one for α_5 . Using the automatic generator [21,23], we obtained a large solver of size 1866×1918 and with 52 solutions.

⁶This solver was not presented in [13] and only the parameterization was discussed.

Our approach: To simplify this solver, we reparameterize the product KR as a function of four new variables, r_1, r_2, r_3 and r_4 . This reparameterization is described in the SM. With this reparameterization, the constraints (22) and (23) give us a set of five equations in five variables r_1, r_2, r_3, r_4 , and α_5 . The two linearly independent equations from (23) are linear in r_1 and r_2 . Hence both these variables can be expressed as functions of the other three variables, r_3, r_4 , and α_5 . Substituting these expressions of r_1 and r_2 in Eq. (22), we obtain a system of three equations in three variables r_3, r_4 , and α_5 , each of degree eight. Let us denote the polynomial set as $E = \{e_1, e_2, e_3\}$. The ideal $I \in \mathbb{C}[r_3, r_4, \alpha_5]$ generated by E is not zero-dimensional [8]. Specifically, if $\alpha_5 = 0$ & $r_3^2 + r_4^2 - 1 = 0$, or $r_3^2 + r_4^2 = 0$, we have a set of trivial solutions to $E = 0$ ⁷.

In order to generate a solver from E , we have to remove those solutions of $E = 0$ where $\alpha_5 = 0$ & $r_3^2 + r_4^2 - 1 = 0$ or $r_3^2 + r_4^2 = 0$. This can be achieved by saturating the ideal I w.r.t. α_5 and $r_3^2 + r_4^2$ [8]. This saturation results to a solver of size 475×501 generated using the generator [21,24].

To generate a smaller solver, we want to avoid saturating w.r.t. α_5 and also avoid adding an extra variable. Thus, we augment E with extra polynomials which vanish on all non-trivial solutions to $E = 0$ but not if $\alpha_5 = 0$ & $r_3^2 + r_4^2 - 1 = 0$. We next show how to generate such polynomials. The form of the i -th polynomial in E is

$$e_i = (r_3^2 + r_4^2 - 1)\phi_i + \alpha_5\psi_i , \quad (24)$$

where ϕ_i and ψ_i are polynomials in r_3, r_4, α_5 . Then, the set of equations $E = 0$ can be written in matrix form as

$$\mathbf{M}\mathbf{b} = \begin{bmatrix} \phi_1 & \psi_1 \\ \phi_2 & \psi_2 \\ \phi_3 & \psi_3 \end{bmatrix}_{3 \times 2} \begin{bmatrix} r_3^2 + r_4^2 - 1 \\ \alpha_5 \end{bmatrix}_{2 \times 1} = \mathbf{0}_{3 \times 1} . \quad (25)$$

The determinant of each 2×2 submatrix of \mathbf{M} vanishes only for those solutions of $E = 0$ such that $\alpha_5 \neq 0$ or $r_3^2 + r_4^2 - 1 \neq 0$. We have a total of three such determinant expressions, each a polynomial of degree ten in the variables r_3, r_4 , and α_5 , out of which two are linearly independent. Let us denote the augmented system consisting of E and two of these polynomials as E_a . This system has 26 solutions and does not contain solutions where $\alpha_5 = 0$ & $r_3^2 + r_4^2 - 1 = 0$. The ideal I_a generated by E_a now has trivial solutions only if $r_3^2 + r_4^2 = 0$. Therefore, we saturate I_a w.r.t. $r_3^2 + r_4^2$ and use Gröbner basis-based method with heuristic sampling [23] to obtain a solver of size 289×315 . We failed to generate a solver using the resultant-based approach [1].

⁷The resulting possible degeneracies for $\alpha_5 = 0$ & $r_3^2 + r_4^2 - 1 = 0$ and $r_3^2 + r_4^2 = 0$ can be avoided by a random rotation of the camera \mathcal{P} .

3.3. H51f

In this scenario, we have five 2D-2D point correspondences $\mathbf{p}_j \leftrightarrow (\mathbf{q}_{ij}, \mathbf{t}_{G_i})$ and one 2D-3D point correspondence $\mathbf{p}_6 \leftrightarrow \mathbf{X}_6^G$, which together give us a set F of 7 equations. For this case we have 5 possible generalized camera configurations, *i.e.*, [1], ..., [5]. We first study the first four configurations.

To solve these configurations we again tested three different groups of parameterizations, *i.e.*, R&t, H, and F together with different simplification, reparameterizations, eliminations and solution strategies. It turns out that the configurations [1], [2], [3], and [4] are the most challenging configurations among all those studied in this work. For these problems the smallest solvers were generated using the **Rotation & translation** parameterization. However, even after different reparameterizations and simplifications these solvers are huge. The sizes of the elimination template matrices for the smallest obtained solvers for these four camera configurations of H51f are listed in Tab. 1. More details about these solvers and parameterizations can be found in the SM.

H51f[5]: In contrast to configurations [1], ..., [4], if all 5 correspondences are detected by the same camera pair, we have the simplest scenario. In this case, the problem is reduced to the problem of estimating the relative pose between two pinhole cameras from 6 2D-2D point correspondences, where the sixth correspondence is obtained by projecting the 3D point into the camera \mathcal{G}_i . This is followed by the scale estimation from the remaining constraint given by the 2D-3D point correspondence. Such a solution was already proposed in [12]. However, there the authors suggested to use the 6pt solver [38] for two-sided common focal length and therefore they reported 15 solutions. This configuration is not practical since \mathcal{G} is usually calibrated and moreover, it usually doesn't have the same focal length as the query camera. In [12], the authors did not test the proposed solver.

In this paper, we consider a more practical scenario where \mathcal{G} is calibrated. This results in estimating the essential matrix and the focal length of \mathcal{P} from six 2D-2D correspondences, known as the one-sided focal length problem [5]. This problem has nine solutions. To solve this problem we use the Gröbner basis-based solver proposed in [20], which eliminates the unknown focal length using the elimination ideal method and thus results in a smaller elimination template matrix of size 9×18 compared to [5]. Using the resultant-based generator [1], we obtained a solver of size 10×10 .

4. Synthetic experiments

For our synthetic scene tests, we generated 5K 3D scenes with known ground truth parameters. In each scene, the

| Solver | KingsCollege | | OldHospital | | ShopFacade | | StMarysChurch | | Runtime |
|------------------------|--------------|-------------|-------------|-------------|-------------|-------------|---------------|-------------|-------------|
| | +LO | | +LO | | +LO | | +LO | | |
| P4Pf | 41.1 | <u>70.3</u> | 33.0 | 52.2 | 79.6 | 89.3 | 59.2 | 83.8 | 0.06 |
| P3.5Pf | 50.4 | 69.4 | 36.8 | <u>55.5</u> | 83.5 | 95.1 | 72.1 | 82.6 | <u>0.08</u> |
| H13f | 61.5 | 69.4 | <u>45.1</u> | 57.7 | 92.2 | 95.1 | <u>76.4</u> | 83.8 | 0.75 |
| H51f[5] | 42.9 | 70.8 | 36.3 | 46.7 | 82.5 | 89.3 | 62.3 | <u>82.8</u> | 0.19 |
| P35Pf + H13f + H51f[5] | <u>60.9</u> | 69.4 | 47.3 | <u>55.5</u> | <u>86.4</u> | <u>90.3</u> | 77.4 | 83.2 | 0.75 |

Table 2. **Hybrid Localization on Cambridge Landmarks.** The table shows the percentage of camera poses localized within 1° and 0.5m. In the table we show the results of both vanilla RANSAC and LO-RANSAC. For a fair comparison the model scoring in RANSAC is the same for all methods (taking into account both 2D-2D and 2D-3D correspondences). The best result is highlighted in bold, and the second-best is underlined. Table also shows the median runtime for RANSAC in seconds.

3D points were randomly distributed within a cube of size $10 \times 10 \times 10$. Each 3D point was projected on up to 6 cameras with realistic focal lengths. Five of these cameras represented the generalized camera \mathcal{G} and one camera was considered as the pinhole camera \mathcal{P} . The orientations and positions of the cameras were selected at random such that they looked towards the origin from a random distance, varying from 15 to 25, from the scene. Images had the resolution of 1000×1000 px. We added Gaussian noise to the positions of the 3D points with the standard deviation σ varying as a % of their depths to simulate the different quality of keypoints used for the triangulation of these 3D points. To simulate the noise in 2D-2D correspondences, we added 2 px image noise. More experiments, *e.g.*, for increasing image noise and fixed 3D point noise are in the SM.

We evaluated the stability of the proposed H32f, H13f and the H51f[5] solvers w.r.t. the SOTA absolute pose solver P4Pf [19]. The graphs for synthetic experiments with increasing noise in the 3D points are provided in Fig. 2. We considered 3 different camera motions in our tests, *i.e.*, random, forward and sideways. For each motion, we measured the error in the estimated rotation \mathbf{R} , translation \mathbf{t} (in SM), and the focal length f , by varying the amount of noise in the 3D points. Note that we tested two derivations of our proposed H13f solver, the one based on the Gröbner basis [23] and one based on resultants [1]. Both solvers have different numerical properties and sizes. Note from Fig. 2, that in the presence of increasing 3D point noise, for all three motions, the Gröbner basis-based H51f[5] solver and the proposed H32f and H13f solvers (both Gröbner basis-based and resultant-based) have much better stability than the SOTA P4Pf absolute pose solver, with the H51f[5] solver slightly outperforming the H32f and H13f solvers. The reason for this behavior is that our hybrid solvers are not only using 2D-3D correspondences, unlike the P4Pf solver. We also observe that the resultant-based solver for H13f has similar or slightly better stability compared to that of the Gröbner basis-based solver for H13f.

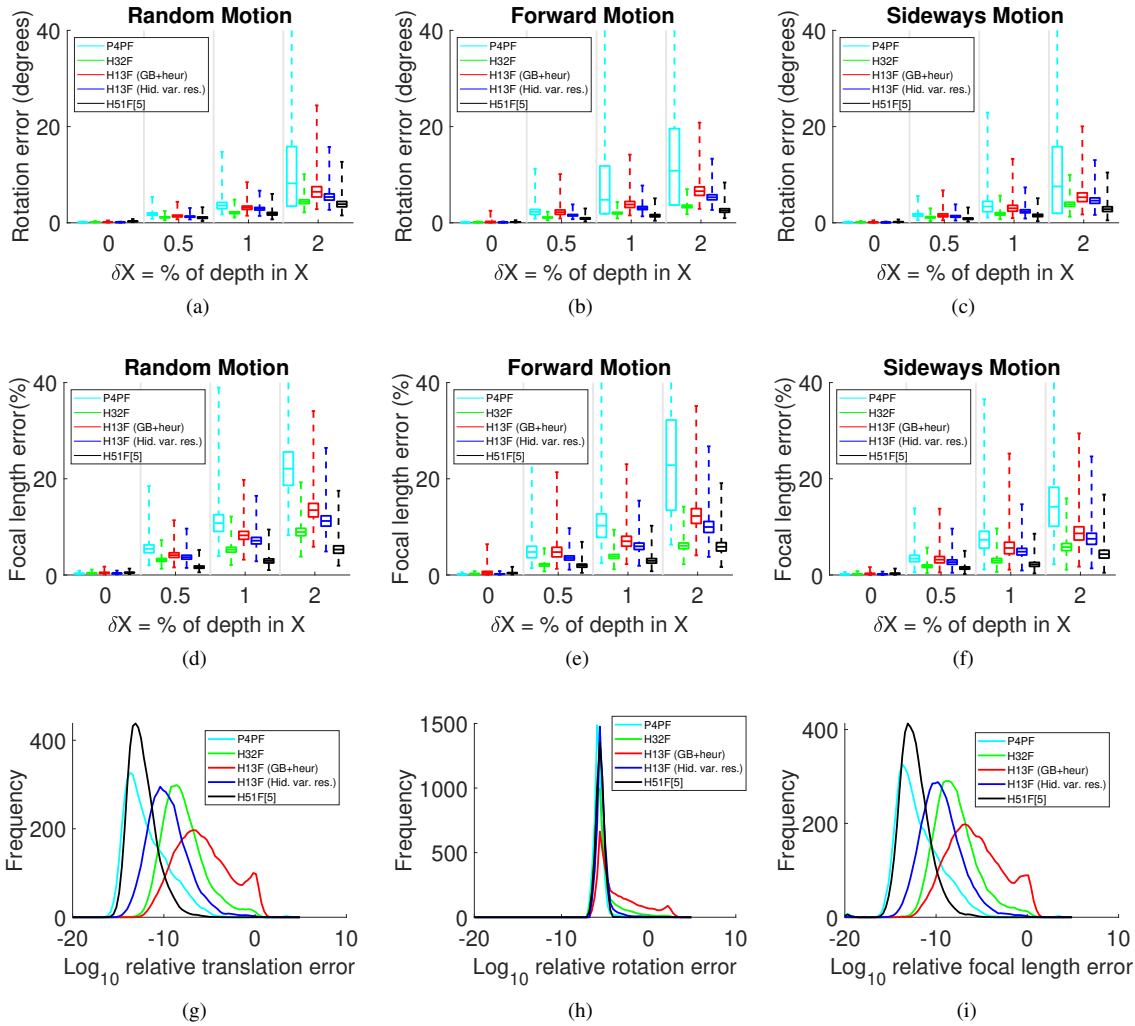


Figure 2. Error in rotation (**Row 1**) and focal length (**Row 2**) in the presence of increasing 3D point noise. Three camera motions considered : random motion (**a,d**), forward motion (**b,e**), and sideways motion (**c,f**). Numerical stability of solvers for noiseless data and random motion (**Row 3**). Our hybrid solvers are more robust to 3D point noise than the **P4Pf** absolute pose solver.

5. Real experiments

We evaluate the proposed **H13f** and **H51f**[5] solvers in a hybrid localization framework and consider four scenes from the Cambridge Landmarks [14] dataset. We did not test our proposed **H32f** solver as its template size makes it less practical for real-time use as compared to the other proposed solvers. For each query image, we establish tentative 2D-2D correspondences to the top-20 retrieved map images based on the DenseVLAD [42] image descriptor. From transitive matching we take all triangulated points as 2D-3D correspondences, and additionally add all 2D-2D correspondences that are either not triangulated, or have a track length less than 5 (as these are potentially less certain 3D points). We apply the solvers in the hybrid LO-RANSAC from [7], minimizing reprojection error (2D-3D)

and Sampson error (2D-2D). We compare with the point-based solvers **P4Pf** [19] and **P3.5Pf** [21], as well as employing all solvers together in the hybrid framework from [7]. Tab. 2 shows the percentage of queries within 1° and 0.5m. To highlight the differences between the accuracy of the solvers, we show the results both with and without local refinement in RANSAC. It can be seen that the proposed **H13f** solver is least affected by noise and returns the most accurate solutions.

6. Conclusion

In this paper, we studied the challenging problem of estimating the semi-genera-lized pose from hybrid point correspondences for partially-calibrated cameras. By testing different parameterizations, elimination techniques and so-

lution strategies, solvers to all minimal configurations of 2D-2D and 2D-3D correspondences, *i.e.*, $\mathbf{H}13f$, $\mathbf{H}32f$ and $\mathbf{H}51f$, and all possible camera configurations within the generalized camera, are derived. The most practical solvers are evaluated on synthetic and real scenes, showing the benefits of hybrid estimation compared to classical 2D-3D approaches. Our solvers fill gaps in the arsenal of minimal solvers and can be used inside hybrid RANSAC [7].

7. Acknowledgements

Torsten Sattler was supported by the EU Horizon 2020 project RICAIP (grant agreement No. 857306) and the European Regional Development Fund under project IMPACT (No. CZ.02.1.01/0.0/0.0/15_003/0000468). Zuzana Kukelova was supported by the OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. Viktor Larsson was supported by the strategic research project ELLIIT.

References

- [1] Snehal Bhayani, Zuzana Kukelova, and Janne Heikkilä. Computing stable resultant-based minimal solvers by hiding a variable. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6104–6111, 2021.
- [2] Snehal Bhayani, Torsten Sattler, Daniel Barath, Patrik Beliansky, Janne Heikkilä, and Zuzana Kukelova. Calibrated and partially calibrated semi-generalized homographies. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5916–5925, 2021.
- [3] Eric Brachmann and Carsten Rother. Expert sample consensus applied to camera re-localization. In *ICCV*, 2019.
- [4] M. Bujnak, Z. Kukelova, and T. Pajdla. A general solution to the p4p problem for camera with unknown focal length. In *CVPR*, 2008.
- [5] Martin Bujnak, Zuzana Kukelova, and Tomas Pajdla. 3d reconstruction from image collections with a single known focal length. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 1803–1810. IEEE, 2009.
- [6] Martin Byrod, Zuzana Kukelova, Klas Josephson, Tomas Pajdla, and Kalle Astrom. Fast and robust numerical solutions to minimal problems for cameras with radial distortion. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [7] Federico Camposeco, Andrea Cohen, Marc Pollefeys, and Torsten Sattler. Hybrid camera pose estimation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 136–144, 2018.
- [8] David A. Cox, John Little, and Donal O’shea. *Using algebraic geometry*, volume 185. Springer Science & Business Media, 2006.
- [9] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. An efficient solution to the homography-based relative pose problem with a common reference direction. In *International Conference on Computer Vision (ICCV)*, 2019.
- [10] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.
- [11] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at <http://www.math.uiuc.edu/Macaulay2/>.
- [12] K. Josephson and M. Byröd. Pose estimation with radial distortion and unknown focal length. In *CVPR*, 2009.
- [13] Klas Josephson, Martin Byrod, Fredrik Kahl, and Kalle Astrom. Image-based localization using hybrid feature correspondences. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
- [14] Alex Kendall, Matthew Grimes, and Roberto Cipolla. PoseNet: A Convolutional Network for Real-Time 6-DOF Camera Relocalization. In *ICCV*, 2015.
- [15] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Automatic generator of minimal problem solvers. In *European Conference on Computer Vision*, pages 302–315. Springer, 2008.
- [16] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Closed-form solutions to minimal absolute pose problems with known vertical direction. In *Asian Conference on Computer Vision*, pages 216–229. Springer, 2010.
- [17] Z. Kukelova, M. Bujnak, and T. Pajdla. Real-Time Solution to the Absolute Pose Problem with Unknown Radial Distortion and Focal Length. In *ICCV*, 2013.
- [18] Zuzana Kukelova, Jan Heller, Martin Bujnak, and Tomas Pajdla. Radial distortion homography. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [19] Zuzana Kukelova, Jan Heller, and Andrew Fitzgibbon. Efficient intersection of three quadrics and applications in computer vision. In *Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [20] Zuzana Kukelova, Joe Kileel, Bernd Sturmfels, and Tomas Pajdla. A clever elimination strategy for efficient minimal solvers. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 4, 2017.
- [21] Viktor Larsson, Kalle Åström, and Magnus Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *CVPR*, volume 2, page 4, 2017.
- [22] Viktor Larsson, Zuzana Kukelova, and Yinqiang Zheng. Making Minimal Solvers for Absolute Pose Estimation Compact and Robust. In *ICCV*, 2017.
- [23] Viktor Larsson, Magnus Oskarsson, Kalle Åström, Alge Wallis, Zuzana Kukelova, and Tomas Pajdla. Beyond gröbner bases: Basis selection for minimal solvers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3945–3954, 2018.
- [24] Viktor Larsson, Kalle Åström, and Magnus Oskarsson. Polynomial solvers for saturated ideals. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2307–2316, 2017.

- [25] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.
- [26] Raul Mur-Artal and Juan D Tardós. ORB-SLAM2: An open-source slam system for monocular, stereo, and RGB-D cameras. *IEEE Transactions on Robotics*, 33(5):1255–1262, 2017.
- [27] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):756–770, 2004.
- [28] Mikael Persson and Klas Nordberg. Lambda twist: An accurate fast robust perspective three point (p3p) solver. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [29] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object Retrieval with Large Vocabularies and Fast Spatial Matching. In *CVPR*, 2007.
- [30] Robert Pless. Camera cluster in motion: motion estimation for generalized camera designs. *IEEE Robotics & Automation Magazine*, 11(4):39–44, 2004.
- [31] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and Marcin Dymczyk. From Coarse to Fine: Robust Hierarchical Localization at Large Scale. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [32] T. Sattler, B. Leibe, and L. Kobbelt. Efficient & Effective Prioritized Matching for Large-Scale Image-Based Localization. *PAMI*, 39(9):1744–1756, 2017.
- [33] Olivier Saurer, Pascal Vasseur, Rémi Boutteau, Cédric Demonceaux, Marc Pollefeys, and Friedrich Fraundorfer. Homography based egomotion estimation with a common direction. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):327–341, 2017.
- [34] Johannes L Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4104–4113, 2016.
- [35] Thomas Schöps, Viktor Larsson, Marc Pollefeys, and Torsten Sattler. Why Having 10,000 Parameters in Your Camera Model Is Better Than Twelve. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [36] Noah Snavely, Steve Seitz, and Richard Szeliski. Photo tourism: exploring photo collections in 3d. In *ACM transactions on graphics (TOG)*, volume 25, pages 835–846. ACM, 2006.
- [37] Eric Stenborg, Torsten Sattler, and Lars Hammarstrand. Using Image Sequences for Long-Term Visual Localization. In *3DV*, 2020.
- [38] Henrik Stewenius, David Nister, Fredrik Kahl, and Frederik Schaffalitzky. A minimal solution for relative pose with unknown focal length. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 789–794. IEEE, 2005.
- [39] Henrik Stewenius, D. Nistér, M. Oskarsson, and K. Åström. Solutions to minimal generalized relative pose problems. 2005.
- [40] Chris Sweeney, Laurent Kneip, Tobias Höllerer, and Matthew Turk. Computing similarity transformations from only image correspondences. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3305–3313, 2015.
- [41] Chris Sweeney, Torsten Sattler, Tobias Hollerer, Matthew Turk, and Marc Pollefeys. Optimizing the viewing graph for structure-from-motion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 801–809, 2015.
- [42] Akihiko Torii, Relja Arandjelović, Josef Sivic, Masatoshi Okutomi, and Tomas Pajdla. 24/7 place recognition by view synthesis. In *CVPR*, 2015.
- [43] Johanna Wald, Torsten Sattler, Stuart Golodetz, Tommaso Cavallari, and Federico Tombari. Beyond Controlled Environments: 3D Camera Re-Localization in Changing Indoor Scenes. In *European Conference on Computer Vision (ECCV)*, 2020.
- [44] T. Weyand and B. Leibe. Discovering Favorite Views of Popular Places with Iconoid Shift. In *ICCV*, 2011.
- [45] Changchang Wu. Towards linear-time incremental structure from motion. In *2013 International Conference on 3D Vision-3DV 2013*, pages 127–134. IEEE, 2013.
- [46] Changchang Wu. P3.5p: Pose estimation with unknown focal length. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2440–2448, 2015.
- [47] Zhengyou Zhang. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22:1330–1334, December 2000.
- [48] Enliang Zheng and Changchang Wu. Structure from motion using structure-less resection. In *International Conference on Computer Vision (ICCV)*, 2015.
- [49] Enliang Zheng and Changchang Wu. Structure From Motion Using Structure-Less Resection. In *The IEEE International Conference on Computer Vision (ICCV)*, 2015.