Three-view Focal Length Recovery From Homographies

Yaqing Ding[1] Viktor Kocur[2] Zuzana Berger Haladová[2] Qianliang Wu[3]
Shen Cai[4] Jian Yang[3] Zuzana Kukelova[1]

[1] Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
{yaqing.ding, kukelzuz}@fel.cvut.cz

[2] Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava
{viktor.kocur, haladova}@fmph.uniba.sk

[3] PCA Lab, Nanjing University of Science and Technology, Nanjing, China
{wuqiangliang, csjyang}@njust.edu.cn

[4] Visual and Geometric Perception Lab, Donghua University
hammer_cai@163.com

Abstract

In this paper, we propose a novel approach for recovering focal lengths from three-view homographies. By examining the consistency of normal vectors between two homographies, we derive new explicit constraints between the focal lengths and homographies using an elimination technique. We demonstrate that three-view homographies provide two additional constraints, enabling the recovery of one or two focal lengths. We discuss four possible cases, including three cameras having an unknown equal focal length, three cameras having two different unknown focal lengths, three cameras where one focal length is known, and the other two cameras have equal or different unknown focal lengths. All the problems can be converted into solving polynomials in one or two unknowns, which can be efficiently solved using Sturm sequence or hidden variable technique. Evaluation using both synthetic and real data shows that the proposed solvers are both faster and more accurate than methods relying on existing two-view solvers. The code and data are available on https://github.com/kocurvik/hf.

1. Introduction

Estimating relative camera motion from multiple views using point correspondences is a classical problem in computer vision. Efficient solutions for various camera configurations have been well-studied in the literature [4, 18–21, 27, 32, 36, 37, 41]. For example, with two fully calibrated cameras, the relative camera pose can be efficiently determined using 5-point algorithms [19, 27, 36]. In contrast to the two view case, the calibrated three-view relative pose problem is much more challenging. This problem can be solved using four triplets of point correspondences [37, 38],[1] however it results in a very complex system of polynomial equations. Thus, the existing solutions [24, 37] to this problem are only approximate and can often fail, *i.e.*, the returned solution can be arbitrarily far from the geometrically correct solution.

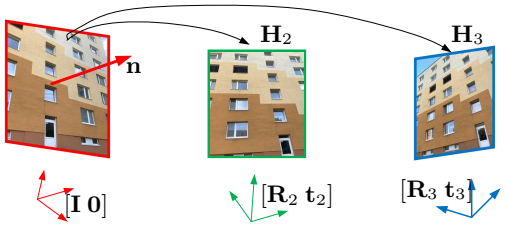


Figure 1. Three cameras view the same plane, defining two homographies $\mathbf{H}_2$, $\mathbf{H}_3$. The two homographies have the same reference image, which should correspond to the same normal vector $\mathbf{n}$.

An important scenario arises when the focal length is the only unknown in the camera intrinsic parameters, referred to as the partially calibrated case. This setup is practical since most modern cameras have zero skew and a centered principal point. When two cameras share an equal yet unknown focal length, their relative motion and the common focal length can be estimated using six point correspondences [19, 27, 28, 41]. Similarly, if one camera is fully calibrated and the other has an unknown focal length, six point correspondences are also required [5, 27, 28]. For

[1] Note, that configuration of four points in three views generates an over-constrained problem. In this case, we have 12 constraint for 11 degrees of freedom (DOF). A minimal solution would need to drop one constraint, *e.g.*, by considering only a line passing through one of the points in the third view [24] or by considering a "half" point correspondence.

two cameras with different, unknown focal lengths, at least seven point correspondences are necessary to recover both the relative motion and the focal lengths [3, 18].

Three-view focal length problems are significantly more complex, often involving polynomial systems with hundreds of solutions. For example, when the three cameras share the same focal length, there can be up to 668 possible solutions [7, 10, 13]. Such systems can be solved using homotopy continuation methods [22, 39]. The running times of the CPU variants of the solvers for four points in three cameras with unknown shared focal length range from $250ms$ to $1456ms$. Efficient GPU implementations are much faster; however, with runtimes $16.7ms$ to $154ms$ they are still too slow for practical applications.

Scenes with planar surfaces, such as floors, walls, doors, facades, and other common structures, are prevalent in man-made environments. When points are coplanar, homography-based algorithms require four point correspondences [20] to estimate the relative pose of two cameras. Unfortunately, any attempt to recover intrinsic parameters from two views of a planar surface (using only point correspondences without additional priors) is futile, as stated in the following theorem [35, 36]: "For any choice of intrinsic parameters, any homography can be realized between two views by positioning the views and a plane in a particular way." However, it has been shown that focal lengths can be recovered from three-view homographies [23, 34], *i.e*. three cameras observing a planar surface (see illustration in Figure 1).

In this paper, we propose novel solutions to the problem of estimating the focal lengths from three-view homographies. We consider four possible cases: (i) three cameras having an unknown equal focal length, (ii) three cameras having two different unknown focal lengths, (iii) three cameras with one known focal length and two unknown equal focal lengths, and (iv) three cameras with one known focal length and two unknown different focal lengths (see Table 1). We propose novel formulations to these problems and discuss the number of possible solutions for each case. The main contributions of the paper include:

- We solve the three-view focal length problems using four coplanar points. We use the property that four coplanar points in three views define two independent homographies, which should share the same plane normal vector.
- Based on the normal vector consistence, we derive new explicit constraints on the focal lengths and the homographies. We provide a detailed problem formulation for the above-mentioned cases, which are then converted into solving polynomial systems in only one or two unknowns.
- This allows us to develop solvers using efficient Sturm sequences or hidden variable technique. The proposed solvers are significantly more efficient than the existing solvers to this problem [23]. Moreover, in extensive synthetic and real experiments, we show an improvement in accuracy over the state-of-the-art two-view solvers.
- We propose a new dataset consisting of 6 scenes (four indoor and 2 outdoor) containing 1870 images captured with 14 different cameras. Ground truth focal lengths are estimated using the standard calibration method [44]. We make both the dataset and the code publicly available.
- To our knowledge, we are the first to extensively evaluate focal length self-calibration from three views of planar scenes on large amounts of synthetic and real scenes, and compare the solutions with different two-view and three-view baselines for cameras observing general scenes.

2. Related Work

Homography estimation is a well-studied problem in the literature and can be solved using the 4-point algorithm, which involves solving a system of eight homogeneous linear equations [20]. However, with only point correspondences, two-view homography does not provide additional constraints on the focal length, making the two-view homography not suitable for self-calibration. To recover the intrinsic parameters from the two-view homography, we need additional priors. In [4], under pure rotation assumption, the authors propose two minimal solutions for the two-view homography-based focal length estimation with one or two focal length parameters. In [15], the authors use sensor-fusion, *i.e*., combining camera with gravity from IMU (Inertial measurement unit), to reduce the DOF of the motion parameters. In general, the gravity prior can provide up to two constraints on the focal length.

Focal length recovery from homographies in more views was first discussed in [34]. Malis *et al*. compute homographies among multiple views, iteratively testing various focal lengths to find the one that minimizes a cost function. However, this method is time-consuming and sensitive to noise.

The work most closely related to ours is by Heikkilä [23], where the constraints on the focal length and the normal vector are derived, and non-iterative solutions to the problem of recovering focal lengths from three-view homographies is proposed. For three cameras with the same unknown focal length, [23] solves three polynomials in three unknowns. However, these polynomials have very high degrees, making them challenging to solve efficiently. Heikkilä first applies SVD to a 82×82 matrix, followed by Gaussian-Jordan elimination of a 82×164 matrix. Finally, the solver finds the solutions by computing the eigenvalues of a 82×82 matrix. For three cameras with two different focal lengths, Heikkilä's method needs to solve four polynomials in four unknowns by performing Gaussian-Jordan elimination of a 176×352 matrix and computing the eigenvalues of a 176×176 matrix.

In contrast to complex solutions presented in [23], we propose very efficient solutions to the problem of recover-

ing focal lengths from three-view homographies. For cameras with the same focal length, our novel solver only requires computing the roots of a univariate polynomial of degree 9, which can be efficiently solved using Sturm sequences. For cameras with two different focal lengths, we compute solutions from the eigenvalues of a 18×18 matrix.

For general scenes, focal lenghts can be recovered together with the epipolar geometry. For two cameras with equal unknown focal lengths, the minimal 6-point solvers [5, 19, 27, 28, 41] use rank-2 and trace constraints on the essential matrix. The 5-point solver [42] solves the plane+parallax scenario for cameras with unknown focal length using four coplanar points and one off-plane point. This solver can be used within the DEGENSAC framework [9] to complement the six point solvers in order to improve performance for scenes that are not completely planar, but contain a dominant plane. For the single unknown focal length problem, the two-view geometry can be solved using the single-side 6-point solver [5, 28]. For the relative pose problem with different and unknown focal lengths, the solution computes the fundamental matrix using the 7-point solver [20], followed by the extraction of the focal lengths from this matrix [3, 25].

As discussed above, three-view focal length problems for general scenes are significantly more complex and their homotopy continuation solutions [7, 10, 13] are impractical.

3. Problem Statement

In this section, we introduce the three-view focal length problem for coplanar points along with the used notation. We consider a set of 3D points $\{\mathbf{X}_i\}$ which lie on a plane defined by

$$\mathbf{n}^\top \mathbf{X}_i = d, \tag{1}$$

where $\mathbf{n} = [n_x, n_y, n_z]$ is the unit normal of the plane and d is the distance of the plane to origin.

The points are observed by three cameras such that $\mathbf{X}_i$ is projected to a 2D point $\mathbf{m}_{i,j}$ by the j-th camera $\mathbf{K}_j[\mathbf{R}_j \mid \mathbf{t}_j]$. In many practical scenarios, it is often reasonable to assume that cameras have square-shaped pixels and that the principal point coincides with the image center [19]. This leaves only the focal length unknown, and thus $\mathbf{K}_j = \mathrm{diag}(f_j, f_j, 1)$.

Without loss of generality, we set the coordinate system such that $\mathbf{R}_1 = \mathbf{I}$ and $\mathbf{t}_1 = \mathbf{0}$. We use $\mathbf{X}_{i,j}$ to denote the points $\{\mathbf{X}_i\}$ expressed in the coordinate system of the j-th camera giving us

$$\mathbf{X}_{i,j} = \mathbf{R}_j \mathbf{X}_i + \mathbf{t}_j, j = 2, 3. \tag{2}$$

Substituting (1) into (2) we obtain

$$\mathbf{X}_{i,j} = \mathbf{R}_j \mathbf{X}_i + \frac{\mathbf{t}_j}{d} \mathbf{n}^\top \mathbf{X}_i = \mathbf{H}_j \mathbf{X}_i, \tag{3}$$

where $\mathbf{H}_j = \mathbf{R}_j + \frac{\mathbf{t}_j}{d}\mathbf{n}^\top$ is the Euclidean homography matrix. Since we have

$$\mathbf{X}_{i,j} \sim \mathbf{K}_j^{-1} \mathbf{m}_{i,j}, \tag{4}$$

further substituting (4) into (3) and expressing $\mathbf{X}_i$ in (3) in the coordinate system of the first camera, we obtain

$$\mathbf{K}_j^{-1} \mathbf{m}_{i,j} \sim \mathbf{H}_j \mathbf{K}_1^{-1} \mathbf{m}_{i,1}, \tag{5}$$

which can be reformulated as

$$\begin{aligned} &[\mathbf{m}_{i,j}]_\times \mathbf{G}_j \mathbf{m}_{i,1} = \mathbf{0}, \\ &\text{with } \mathbf{G}_j \sim \mathbf{K}_j \mathbf{H}_j \mathbf{K}_1^{-1}, \end{aligned} \tag{6}$$

where $\sim$ indicates equality up to a scale factor and $\mathbf{G}_j$ represents homography between the first and the j^{th} camera.

We can use equation (6) to obtain 2D homographies $\mathbf{G}_j$, $j = 2, 3$ from point correspondences between the first, and the second and the first and the third camera. Now, our goal is to estimate the unknown focal lengths f_j from $\mathbf{G}_2$ and $\mathbf{G}_3$. There are several possible configurations of three cameras based on combinations of known and unknown equal or different focal lengths. In Section 5 we derive solvers for four such camera configurations denoted as Cases I-IV, which we list in Table 1.

We derive the new solvers by utilizing a key observation that both $\mathbf{H}_2$ and $\mathbf{H}_3$ are related to the normal vector $\mathbf{n}$. In the next section, we use this to show that $\mathbf{G}_2$ and $\mathbf{G}_3$ can provide two constraints on the intrinsic parameters, which can be used to recover the focal lengths.

4. Our Approach

Given two 2D homographies $\mathbf{G}_j$, we have

$$\mathbf{H}_j \sim \mathbf{K}_j^{-1} \mathbf{G}_j \mathbf{K}_1. \tag{7}$$

Thus for known $\mathbf{G}_j$, $\mathbf{H}_j$ are polynomial matrices in the focal length parameters. To solve for the focal lengths, we need to find the constraints on $\mathbf{H}_j$. Based on the formulation of the Euclidean homography (3), we know that [20]

$$\mathbf{E}_j = [\mathbf{t}_j]_\times \mathbf{H}_j = [\mathbf{t}_j]_\times (\mathbf{R}_j + \frac{\mathbf{t}_j}{d}\mathbf{n}^\top) = [\mathbf{t}_j]_\times \mathbf{R}_j, \tag{8}$$

where $\mathbf{E}_j$ is the essential matrix corresponding to $\mathbf{H}_j$. Let's consider $\mathbf{H}_j^\top$, where

$$\mathbf{H}_j^\top = \mathbf{R}_j^\top + \frac{\mathbf{n}}{d}\mathbf{t}_j^\top. \tag{9}$$

It can be easily shown that the essential matrix $\tilde{\mathbf{E}}_j$ corresponding to $\mathbf{H}_j^\top$ is given by

$$\tilde{\mathbf{E}}_j = [\mathbf{n}]_\times \mathbf{H}_j^\top = [\mathbf{n}]_\times (\mathbf{R}_j^\top + \frac{\mathbf{n}}{d}\mathbf{t}_j^\top) = [\mathbf{n}]_\times \mathbf{R}_j^\top. \tag{10}$$

Problems	Proposed Solvers	Focal length			No. of Solutions		Eigenvalue		Time (μs)	
		View 1	View 2	View 3	Ours	Heikkilä [23]	Ours	Heikkilä [23]	Ours	Heikkilä [23]
Case I	$\mathbf{H}_{fff}$	f	f	f	9	70	Sturm	82×82	17.3	1404
Case II	$\mathbf{H}_{ff}$	Known	f	f	6	-	Sturm	-	19.4	-
Case III	$\mathbf{H}_{f\rho\rho}$	f	ρ	ρ	17	152	18×18	176×176	200	5486
Case IV	$\mathbf{H}_{f\rho}$	Known	f	ρ	9	-	12×12	-	106	-

Table 1. Four possible cases for three views. i) Equal and unknown focal length for three cameras; ii) Known focal length of reference camera, equal and unknown focal length for target cameras; iii) Focal lengths of reference camera and two target cameras are different; iv) Known focal length of reference camera, different and unknown focal lengths for target cameras.

Thus the essential matrices derived from different $\mathbf{H}_j^\top$ are related by the same normal vector $\mathbf{n}$ in the reference camera coordinate. As shown in [36, 40], a valid essential matrix should satisfy the singular and trace constraints

$$\begin{aligned} \det(\tilde{\mathbf{E}}_j) &= 0, \\ \tilde{\mathbf{E}}_j \tilde{\mathbf{E}}_j^\top \tilde{\mathbf{E}}_j - \frac{1}{2}\mathrm{trace}(\tilde{\mathbf{E}}_j \tilde{\mathbf{E}}_j^\top)\tilde{\mathbf{E}}_j &= \mathbf{0}. \end{aligned} \tag{11}$$

In our case, we omit the zero determinant constraint since $\tilde{\mathbf{E}}_j$ is already singular by construction (10).

Substituting (7) into the trace constraints (11) gives us nine equations per homography, *i.e.*, 18 equations for two homography matrices $\mathbf{H}_2$ and $\mathbf{H}_3$. Only two from the nine equations are algebraically independent, *i.e.*, one homography matrix provides two constraints on the focal lengths and the normal vector parameters. For the case of equal unknown focal length, we have only 3 unknowns (f, n_x, n_y) (since (11) is homogenous in $\mathbf{n}$, we can let $n_z = 1$) and thus an over-constrained problem. The problem can be solved using only one of the nine trace constraint equations for $\tilde{\mathbf{E}}_2$. These equations can be solved using the Gröbner basis method [29, 31] The final solver performs Gauss-Jordan elimination of a 118×141 matrix and extracts solutions from the eigenvectors of a 23×23 matrix. However, this solver is not efficient enough for real-time applications.

To derive more efficient solutions, we use additional constraints. It can be seen that

$$\begin{aligned} [\mathbf{n}]_\times \mathbf{H}_j^\top \mathbf{H}_j [\mathbf{n}]_\times^\top &= [\mathbf{n}]_\times \mathbf{R}_j^\top \mathbf{R}_j [\mathbf{n}]_\times^\top, \\ &= [\mathbf{n}]_\times [\mathbf{n}]_\times^\top. \end{aligned} \tag{12}$$

Substituting (7) into (12) gives

$$[\mathbf{n}]_\times \mathbf{Q}_j [\mathbf{n}]_\times^\top \sim [\mathbf{n}]_\times [\mathbf{n}]_\times^\top, \tag{13}$$

with

$$\mathbf{Q}_j = (\mathbf{K}_j^{-1}\mathbf{G}_j\mathbf{K}_1)^\top (\mathbf{K}_j^{-1}\mathbf{G}_j\mathbf{K}_1). \tag{14}$$

Note that, we use $\sim$ instead of equality in (13) since $\mathbf{H}_j$ from (7) is up to a scale factor. Since $\mathbf{Q}_j$ are symmetric matrices, we can write them as

$$\mathbf{Q}_j = \begin{bmatrix} q_{j1} & q_{j2} & q_{j3} \\ q_{j2} & q_{j4} & q_{j5} \\ q_{j3} & q_{j5} & q_{j6} \end{bmatrix}. \tag{15}$$

Then (13) can be rewritten as

$$[\mathbf{n}]_\times \mathbf{Q}_j [\mathbf{n}]_\times^\top = s_j [\mathbf{n}]_\times [\mathbf{n}]_\times^\top, \; j = 2, 3 \tag{16}$$

where we add scale factors to ensure the equality. Note that both the left and right of (16) are symmetric matrices, hence we can get 6 equations for each j.

To simplify our 12 equations, we can eliminate some unknowns from these equations using the elimination ideal technique [11]. This technique was recently used to solve several minimal camera geometry problems [28].

In our case, we first create an ideal J generated by 12 polynomials (16). Then, the unknown elements of the normal $\{n_x, n_y\}$ and the scale factor s_2, s_3 are eliminated from the generators of J by computing the generators of the elimination ideal $J_1 = J \cap \mathbb{C}[q_{21}, \ldots, q_{36}]$. Here, $q_{j\cdot}$ are the entries of $\mathbf{Q}_2, \mathbf{Q}_3$. These generators can be computed using computer algebra software Macaulay2 [17] (for more details and the input Macaulay2 code see the SM).
In this case, the elimination ideal J is generated by seven polynomials g_i of degree 6 in the elements of $\mathbf{Q}_j$, $j = 2, 3$. The final constraints are only related to the 12 elements of the symmetric matrices $\mathbf{Q}_i$ (6 from $\mathbf{Q}_2$ and 6 from $\mathbf{Q}_3$). To the best of our knowledge, these constraints are first shown in this paper and have not been used in the computer vision literature before. Alternatively, we can eliminate n_x, n_y from the 18 equations of the trace constraints (11), however, this will result in more complicated equations since the constraints will be related to the 18 elements of $\mathbf{H}_i$.

5. New Solvers

In this section, we propose solvers for the four different cases outlined in Tab. 1 using the constraints derived in the previous section. We also propose a method based on LO-RANSAC [8] that utilizes the new solvers for robust estimation of focal lengths from three views of planar scenes.

5.1. One Unknown Focal Length Parameter

Case I. We first consider the case where the three cameras have equal and unknown focal length, *i.e.*, $f_1 = f_2 = f_3 = f$, and $\mathbf{K}_{1,2,3} = \mathrm{diag}(f, f, 1)$. By substituting $\mathbf{K}_j$ into the generators g_i, $i = 1, \ldots, 7$ of the elimination ideal J, we

obtain 7 univariate polynomials in f, which are of degree 9 in $\alpha = f^2$. They form an over-constrained system. We only need one of them to find the solutions to f. To find the roots of the degree 9 univariate polynomial we use the Sturm sequence method [16]. We denote this solver as $\mathbf{H}_{fff}$.

Case II. In the second case, we assume that f_1 is known, and $f_2 = f_3 = f$ are unknown. This case occurs when a calibrated camera is used to capture the reference image, and the second uncalibrated camera captures two target images. Similar to Case I, we only need to find the roots of a univariate polynomial in f, in this case of degree 6, using Sturm sequences [16]. We denote this solver as $\mathbf{H}_{ff}$.

5.2. Two Unknown Focal Length Parameters

In the second group of solvers, we consider self-calibration problems with two different unknown focal lengths. There are two practical cases:

Case III. In this case, we assume two unknown focal lengths $f_1 = f$, and $f_2 = f_3 = \rho$, *i.e.*, $\mathbf{K}_1 = \mathrm{diag}(f, f, 1)$ and $\mathbf{K}_{2,3} = \mathrm{diag}(\rho, \rho, 1)$. This corresponds to a situation where the first uncalibrated camera is used to capture the reference image and the second uncalibrated camera captures two target images. By substituting $\mathbf{K}_j$ into the constraints g_i from the elimination ideal, we obtain 7 polynomials in two unknowns $\alpha, \beta, (\alpha = f^2, \beta = \rho^2)$, which can be written as

$$\mathbf{M}\mathbf{v} = \mathbf{0}, \tag{17}$$

where $\mathbf{M}$ is a 7×28 coefficient matrix and

$$\mathbf{v} = [1, \beta, ..., \beta^6, \alpha, \alpha\beta, ..., \alpha\beta^6, ..., \alpha^3, ..., \alpha^3\beta^6]^\top, \tag{18}$$

is a vector consisting of the 28 monomials. The system of polynomial equations in (17) can be solved using different algebraic methods [11]. There are also different state-of-the-art approaches for generating efficient algebraic solvers [2, 19, 26, 27, 31]. In this paper, we use the hidden variable technique to derive polynomial eigenvalue solution based on [27].

Polynomial Eigenvalue Solution. The polynomial system in (17) contains four polynomials in two unknowns (α, β), and the highest degree of the unknown α is 3. In this case, α can be chosen as the hidden variable, *i.e.* we can consider it as a parameter. Then the system of polynomial equations (17) can be rewritten as

$$\mathbf{C}(\alpha)\tilde{\mathbf{v}} = \mathbf{0}, \tag{19}$$

where $\mathbf{C}(\alpha)$ is a 7×7 polynomial matrix parameterized by α, and $\tilde{\mathbf{v}} = [1, \beta, ..., \beta^6]^\top$ is a vector of monomials in β without α. $\mathbf{C}(\alpha)$ can be rewritten as

$$\mathbf{C}(\alpha) = \alpha^3\mathbf{C}_3 + \alpha^2\mathbf{C}_2 + \alpha\mathbf{C}_1 + \mathbf{C}_0, \tag{20}$$

where $\mathbf{C}_3, \mathbf{C}_2, \mathbf{C}_1, \mathbf{C}_0$ are 7×7 coefficient matrices containing only numbers. For this problem, the matrix $\mathbf{C}_3$ is only rank 4, resulting in four zero eigenvalues. To speed up the solver, we remove these zero eigenvalues from the computations. To do this, we first need to transform the matrices $\mathbf{C}_i$ by considering linear combinations of their rows, such that there are three zero rows in the transformed matrix $\mathbf{C}_3$. To remove the zero rows in the transformed $\mathbf{C}_3$, we use the technique from [15] and consider the transpose of (19)

$$\tilde{\mathbf{v}}^\top\mathbf{C}^\top(\alpha) = \mathbf{0}. \tag{21}$$

In this case, we have

$$\mathbf{C}^\top(\alpha) = \alpha^3\mathbf{C}_3^\top + \alpha^2\mathbf{C}_2^\top + \alpha\mathbf{C}_1^\top + \mathbf{C}_0^\top. \tag{22}$$

The zero rows in $\mathbf{C}_3$ are now zero columns in $\mathbf{C}_3^\top$. Since $\mathbf{C}_0$ is full rank, we let $\gamma = \frac{1}{\alpha}$ and rewrite (22) as

$$\mathbf{C}^\top(\gamma) = \gamma^3\mathbf{C}_0^\top + \gamma^2\mathbf{C}_1^\top + \gamma\mathbf{C}_2^\top + \mathbf{C}_3^\top. \tag{23}$$

If we consider (23) as a polynomial eigenvalue problem [1], the solutions to γ are the eigenvalues of 21×21 matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \\ -\mathbf{C}_0^{-\top}\mathbf{C}_3^\top & -\mathbf{C}_0^{-\top}\mathbf{C}_2^\top & -\mathbf{C}_0^{-\top}\mathbf{C}_1^\top \end{bmatrix}. \tag{24}$$

The three zero columns in $\mathbf{C}_3^\top$ can now be removed together with their corresponding rows to eliminate the zero eigenvalues [27]. In this way, we obtain 18 possible solutions. Once we have solutions to α, the remaining unknown β can be extracted from the null vector of $\mathbf{C}(\alpha)$ based on (19). We denote this solver as $\mathbf{H}_{f\rho\rho}$.

Note that in the proposed polynomial eigenvalue formulation, we solve a relaxed version of the original problem (11). The original system (11) has 17 solutions, as it can be shown, *e.g.*, using the computer algebra system Macaulay2 [17]. In the polynomial eigenvalue formulation, we have one spurious solution that does not ensure that the elements of $\tilde{\mathbf{v}}$ satisfy $\tilde{\mathbf{v}} = [1, \beta, \cdots, \beta^6]^\top$.

Case IV. Finally, we consider a case where f_1 is known and $f_2 = f$, $f_3 = \rho$ are unknown. The solver for this case, we denote it as $\mathbf{H}_{f\rho}$, performs steps similar to the solver $\mathbf{H}_{f\rho\rho}$ for Case III. In this case, the $\mathbf{H}_{f\rho}$ solver computes the eigenvalues of a 12×12 matrix. Due to space limitations, we describe the $\mathbf{H}_{f\rho}$ solver for Case IV in the SM.

5.3. Robust Estimation of Focal Lengths

To estimate focal lengths from three images of planar scenes we utilize the LO-RANSAC framework [8] using a strategy inspired by [36]. We first extract triplet point correspondences from images (e.g. using [12, 33]). In each RANSAC iteration we sample 4 triplets from which we estimate $\mathbf{G}_2$ and $\mathbf{G}_3$ using DLT [20]. These matrices are used as inputs to solvers proposed in Section 5.1 and 5.2.

For each resulting real positive solution we use (7) to obtain $\mathbf{H}_2$, which is then decomposed into two poses $(\mathbf{R}_2, \mathbf{t}_2)$.[2] We then use these poses and focal lengths to triangulate three of the sampled points in the first two views thus obtaining points $\mathbf{X}_{i1}$. We use the corresponding points in the third view to obtain $\mathbf{R}_3$ and $\mathbf{t}_3$ using the $\mathbf{P3P}$ solver [14]. Note that it is possible to obtain $\mathbf{R}_3$ and $\mathbf{t}_3$ by decomposing $\mathbf{H}_3$ and using a single correspondence to obtain the scale of $\mathbf{t}_3$, but in practical experiments we found the approach using $\mathbf{P3P}$ faster and more accurate. We score the generated models using pairwise Sampson error for each of the three image pairs. Whenever a new so-far-best model is found we perform local optimization using the Levenberg-Marquardt algorithm minimizing the pairwise Sampson error across the three pairs of views. In addition to the estimated focal lengths this strategy also produces the relative poses of all three cameras. We denote the proposed robust estimators as $\mathbf{H}_{fff} + \mathbf{P3P}$ for Case I, $\mathbf{H}_{ff} + \mathbf{P3P}$ for Case II, $\mathbf{H}_{f\rho\rho} + \mathbf{P3P}$ for Case III and $\mathbf{H}_{f\rho} + \mathbf{P3P}$ for Case IV to distinguish them from the solvers. Our implementation is based on PoseLib [30].

For Case I and II this approach is able to find the single unknown focal length even when the scene is fully planar. For Case III and IV we may obtain multiple solutions from which the correct focal lengths can not be distinguished using only planar points thus requiring some off-plane points during scoring and local optimization.

6. Experiments

We perform extensive experiments on synthetic and real data to evaluate the performance of the proposed solvers and the robust focal length estimation strategy described in Section 5.3. We compare our solvers with several baselines using either pairwise or triplet correspondences. For pairwise correspondences, we consider the 6 point solver for relative pose with one unknown shared focal length [27] denoted as $f\mathbf{E}f$and its combination with the $4+1$ point plane and parallax solver [42] using the DEGENSAC framework [9], which we denote as $f\mathbf{E}f + \mathbf{PP}$. We also evaluate the solver for one unknown focal length [5] denoted as $\mathbf{E}f$. When considering triplets, we use a strategy [36] of using a pairwise solver followed by triangulation of points and registration of the third camera using the $\mathbf{P3P}$ solver [14]. (see Section 5.3 for more details). We denote the methods that work with triplet correspondences as $f\mathbf{E}f + \mathbf{P3P}$, $f\mathbf{E}f + \mathbf{PP} + \mathbf{P3P}$, $\mathbf{E}f + \mathbf{P3P}$ respectively.

Note that we do not compare with the solvers proposed in [23]. As visible from Table 1 these solvers are significantly slower than our solvers [3] and they return significantly more solutions that need to be tested inside RANSAC. As such they are not practical for real-world applications. Moreover, the solver for Case IV proposed in [23] returns solutions only to one of the two unknown focal lengths.

[2]Four poses are possible, but we select those two for which the plane intersects the principal axis in front of the first camera.

[3]We report runtimes of the original Matlab implementations of [23] and the Matlab + mex implementations of our solvers on I7-11700K CPU.

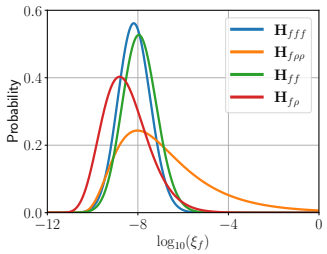


Figure 2. Numerical stability of the proposed solvers.

Except for the numerical stability experiment, due to space constraints, in the main paper we present only results for Case I and Case II. Experiments for Case III and IV for synthetic and real data are in the SM.

6.1. Synthetic experiments

Numerical Stability. We evaluate the numerical stability of the solvers proposed in Sections 5.1 and 5.2. The synthetic data is generated in the following setup. We sample 200 3D points distributed on a plane with a random orientation. The focal lengths of the cameras are generated from a uniform distribution $f_g \in [300, 3000]$ px with a field of view of 90 degrees. The baseline between consecutive cameras is set to be 10 percent of the average scene distance. We generated 10,000 random scenes with 3D points on different planes and different transformations between consecutive views. The focal length error ξ_f is defined as:

$$\xi_f = \frac{|f_e - f_g|}{f_g}, \tag{25}$$

where f_g, f_e represent the ground-truth focal length and the estimated focal length, respectively. For the solvers with different focal lengths, we use the geometric mean of the two focal length errors $\xi_f = \sqrt{\xi_{f_1}\xi_{f_2}}$.

Fig. 2 shows the results of the $\log_{10}$ relative focal length error for the proposed methods by considering the solution closest to the ground truth. As can be seen, all of our solvers are numerically stable without large errors.

Accuracy of the Estimated Focal Lengths. Next, we evaluate the performance of the robust estimators proposed in Section 5.3 on synthetic data. We show how varying noise levels and the proportion of points that lie on a plane affect their accuracy compared to baselines.

The synthetic data is generated with $n = 200$ 3D points visible by the three 1920px $\times$ 1080px cameras with focal lengths sampled unfiormly $f_g \in [300, 3000]$. The cameras are positioned randomly such that their views overlap. n_p of the total n points lie on a plane with random orientation. From the 3D points we create triplets with 75% inlier ratio

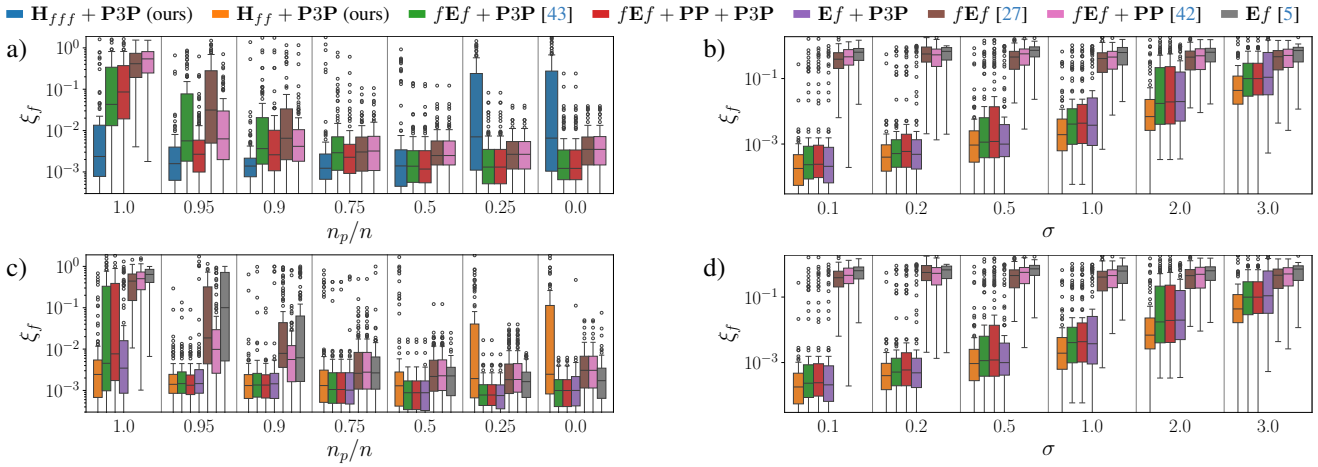


Figure 3. Focal length errors for the evaluated methods in synthetic experiments. **Case I**: (a) We vary the proportion of points which lie on the dominant plane with fixed noise $\sigma = 1$. (b) We vary noise σ with $n_p/n = 1.0$ (*i.e.* all points lie on a plane). **Case II**: (c) We vary the proportion of points which lie on the dominant plane with fixed noise $\sigma = 1$. (d) We vary noise σ with $n_p/n = 1.0$.

	Method		Sample	Median ξ_f	Mean ξ_f	mAA$_f$(0.1)	mAA$_f$(0.2)	Runtime (ms)
Case I	$\mathbf{H}_{fff}+\mathbf{P3P}$	**ours**	4 triplets	**0.0439**	**0.1342**	**51.49**	**65.58**	49.31
	$\mathbf{H}_{fff}+\mathbf{P3P}$	[23]	4 triplets	0.0480	0.1693	48.73	61.94	1102.48
	$f\mathbf{E}f+\mathbf{P3P}$	[43]	6 triplets	0.0503	0.1604	47.83	61.45	37.44
	$f\mathbf{E}f+\mathbf{PP}+\mathbf{P3P}$		6 triplets	0.0518	0.1661	47.11	60.65	41.18
	$f\mathbf{E}f$	[27]	6 pairs	0.4793	0.4924	10.49	17.07	**24.54**
	$f\mathbf{E}f+\mathbf{PP}$	[42]	6 pairs	0.5166	0.5149	10.31	16.51	31.95
Case II	$\mathbf{H}_{ff}+\mathbf{P3P}$	**ours**	4 triplets	**0.0611**	**0.2073**	**43.11**	**55.54**	35.97
	$f\mathbf{E}f+\mathbf{P3P}$	[43]	6 triplets	0.0692	0.2250	40.42	52.35	31.81
	$f\mathbf{E}f+\mathbf{PP}+\mathbf{P3P}$		6 triplets	0.0714	0.2317	39.86	51.58	34.35
	$\mathbf{E}f+\mathbf{P3P}$		6 triplets	0.0691	0.2406	40.60	52.13	36.82
	$f\mathbf{E}f$	[27]	6 pairs	0.4623	0.4781	12.01	19.12	28.59
	$f\mathbf{E}f+\mathbf{PP}$	[42]	6 pairs	0.4968	0.5006	12.11	18.98	37.60
	$\mathbf{E}f$	[5]	6 pairs	0.6827	0.5932	8.87	13.01	**22.59**

Table 2. Comparison of the focal length estimation accuracy of the evaluated methods on our dataset of planar scenes.

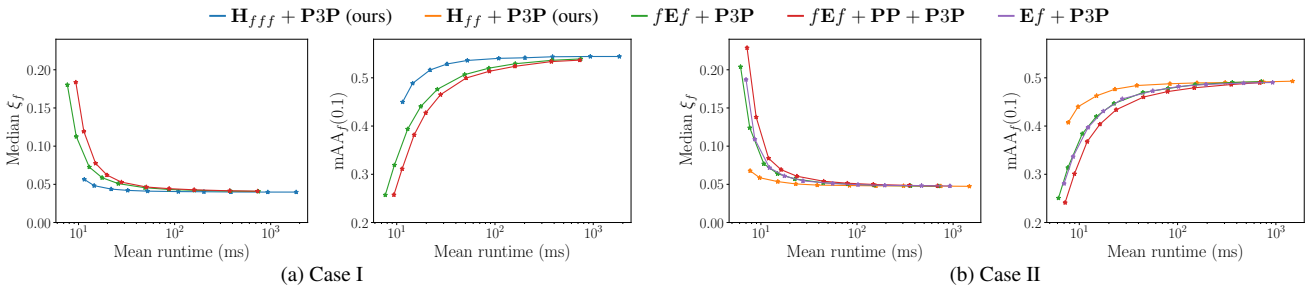
by randomly shuffling 25% of the points. We add Gaussian noise with standard deviation of σ to the pixel coordinates of points. We run all of the tested methods in PoseLib [30] for a fixed number of 100 iterations with the Sampson error threshold of 3 px. For methods that use only pairwise correspondences we consider only the first two views. For each configuration we generate 100 random scenes.

Fig. 3 shows the results of the synthetic experiments in which we vary both σ and n_p. When considering Case I, $\mathbf{H}_{fff}+\mathbf{P3P}$ shows significantly better accuracy than the remaining methods when a large portion of points lie on a single dominant plane (Fig. 3 (a)) even under high levels of noise (Fig. 3 (b)). This indicates a clear superiority of $\mathbf{H}_{fff}+\mathbf{P3P}$ when dealing with images of planar scenes. Similarly, for Case II, $\mathbf{H}_{ff}+\mathbf{P3P}$ also shows improved focal length estimation accuracy when considering planar scenes, although the improvement over competing methods is not as significant. For fully planar scenes, as expected, the two-view solvers $f\mathbf{E}f$, $f\mathbf{E}f+\mathbf{PP}$ and $\mathbf{E}f$ fail to generate correct solutions (Fig. 3 (b),(d)). For scenes with few of the plane points, *e.g.*, scenes with $n_p/n \in [0.95, 0.9]$, the $f\mathbf{E}f+\mathbf{PP}$ solver, *i.e.*, the solver where the $4+1$ plane+parallax solver is utilized within DEGENSAC with $f\mathbf{E}f$ [42], outperforms the $f\mathbf{E}f$ solver (Fig. 3 (a)). This is not surprising since the $f\mathbf{E}f+\mathbf{PP}$ solver [42] was developed to handle scenes with dominant planes and some off-the plane points. The solvers that work with triplet correspondences, *i.e.* $f\mathbf{E}f+\mathbf{P3P}$, $f\mathbf{E}f+\mathbf{PP}+\mathbf{P3P}$, $\mathbf{E}f+\mathbf{P3P}$, outperform the two-view solvers. Additional synthetic experiments are presented in SM.

6.2. Real-World Experiments

Dataset. Since there are no suitable datasets for evaluating focal length recovery on planar scenes, we have collected a new dataset for evaluation which we make publicly available. We have collected the dataset using 14 different cameras. For each camera we obtained ground truth focal lengths using the standard calibration method [44] with a checkerboard pattern. Using these cameras we have captured four indoor and two outdoor planar scenes from various positions resulting 1870 images in total. Sample images of the scenes are shown in Fig. 5. To obtain triplet correspondences we used SuperPoint features [12] with LightGlue matches [33]. More details are available in SM.

For Case I we have randomly sampled 500 image triplets per camera per scene. For Case II/III we sampled 50 triplets



(a) Case I (b) Case II

Figure 4. Median ξ_f and mAA$_f(0.1)$ plotted for different number of RANSAC iterations (10, 20, 50, 100, 200, 500, 1000, 2000, 5000, 10000). All methods are evaluated in combination with **P3P** [14] and non-linear optimization within PoseLib [30]. We do not include methods relying solely on pairwise correspondences since they all result in very low accuracy (median $\xi_f > 0.4$, mAA$_f(0.1) < 0.2$).

Asphalt Boats Book

Facade Floor Papers

Figure 5. The six planar scenes captured in our evaluation dataset.

per camera pair and for Case IV 10 triplets per camera triplet. We only kept image triplets that had at least 10 triplet matches. For some scenes and cameras, it was impossible to obtain 500, 50 or 10 triplets, respectively, resulting in 35 472 image triplets for Case I, 18 219 triplets for Case II/III and 12 876 for Case IV. Details on the cameras, number of images, and triplets are provided in SM.

Evaluation. We evaluate the methods using ξ_f (25) and the mean average accuracy derived from ξ_f [25], denoted as mAA$_f(t)$ representing the normalized area under the curve of the cumulative distribution function of ξ_f on $[0, t]$.

Table 2 shows the results for our $\mathbf{H}_{fff}$ and $\mathbf{H}_{ff}$ solvers and the baselines for RANSAC with a fixed maximum number of 1000 iterations, with early termination at 0.9999 success probability (allowed after the first 100 iterations) and an epipolar threshold of 3px. The results show a significantly better accuracy of $\mathbf{H}_{fff}$ + **P3P** and $\mathbf{H}_{ff}$ + **P3P** compared to the baseline methods. We also include a version of $\mathbf{H}_{fff}$ + **P3P** with our solver replaced with the solver from [23] which we re-implemented from Matlab into C++. We show that the solver from [23] results in worse accuracy than utilizing our novel solver, possibly due to worse numerical stability. In addition, our $\mathbf{H}_{fff}$ solver results in ∼20x faster runtime than [23].

To further demonstrate the efficacy of our method, we also performed speed-vs-accuracy tradeoff evaluation by running each method for a different number of fixed RANSAC iterations. We measured runtime for all methods on one core of Intel Icelake 6338 2GHz processor. The results for our methods and the baselines are shown in Fig. 4. Since our methods use the same non-linear optimization strategy as the baselines which utilize triplets, given enough iterations, the different methods eventually converge to similar focal lengths resulting in similar accuracy. However, the results show that our solvers lead to very accurate results in fewer iterations compared to the baselines, showcasing their practical viability for focal length estimation.

Limitations. Our solvers have several limitations. First, pure translation is a degenerate case for the $\mathbf{H}_{fff}$ and $\mathbf{H}_{f\rho\rho}$ solvers. However, if the focal length of the reference camera is known, we can still recover the focal lengths of the target cameras under pure translation. Due to the outliers, the focal length solvers are usually used in a robust estimation framework such as RANSAC or simpler robust voting framework [6]. In our case, we found that robust voting is not suitable and may fail due to the fact that more than one solution generated by the solvers may be geometrically feasible. Moreover, as mentioned above for Case III and IV focal lengths can not be distinguished using only planar points thus requiring some off-plane points during scoring. Note that this is a property of the problem and not the solvers. Problem with recovering one focal length for Case III was mentioned also in [23].

7. Conclusion

We address the problem of estimating the focal lengths of three cameras observing a planar scene. We derive novel constrains for this problem and use them to propose four new efficient solvers for different possible camera configurations. We extensively evaluate the proposed solvers on both real and synthetic data showing their superiority over baseline approaches. To the best of our knowledge, we are the first to perform such extensive evaluation for these problems. For this purpose, we introduce a new public dataset of planar scenes captured by multiple calibrated cameras.

Acknowledgments

This work was funded by the Czech Science Foundation (GAČR) JUNIOR STAR Grant No. 22-23183M (supporting Y.D. and Z.K.), the TERAIS project, a Horizon-Widera-2021 program of the European Union under the Grant agreement number 101079338 (supporting V.K. and Z.B.H.) and by Slovak Research and Development Agency under project APVV-23-0250, the grant KEGA 004UK-4/2024 "DICH: Digitalization of Cultural Heritage" (supporting Z.B.H.). Q.W. and J.Y. were supported by the National Science Fund of China under Grant Nos. U24A20330 and 62361166670. The results were obtained using the computational resources procured in the project National competence centre for high performance computing (project code: 311070AKF2) funded by European Regional Development Fund, EU Structural Funds Informatization of society, Operational Program Integrated Infrastructure.

References

[1] Zhaojun Bai, James Demmel, Jack Dongarra, Axel Ruhe, and Henk van der Vorst. *Templates for the solution of algebraic eigenvalue problems: a practical guide*. SIAM, 2000. 5

[2] Snehal Bhayani, Zuzana Kukelova, and Janne Heikkila. A sparse resultant based method for efficient minimal solvers. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1770–1779, 2020. 5

[3] Sylvain Bougnoux. From projective to euclidean space under any practical situation, a criticism of self-calibration. In *Sixth International Conference on Computer Vision*, pages 790–796. IEEE, 1998. 2, 3

[4] Matthew Brown, Richard I Hartley, and David Nistér. Minimal solutions for panoramic stitching. In *Computer Vision and Pattern Recognition (CVPR)*, 2007. 1, 2

[5] Martin Bujnak, Zuzana Kukelova, and Tomas Pajdla. 3d reconstruction from image collections with a single known focal length. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1803–1810. IEEE, 2009. 1, 3, 6, 7

[6] Martin Bujnak, Zuzana Kukelova, and Tomas Pajdla. Robust focal length estimation by voting in multi-view scene reconstruction. In *Asian Conference on Computer Vision*, pages 13–24. Springer, 2009. 8

[7] Chiang-Heng Chien, Hongyi Fan, Ahmad Abdelfattah, Elias Tsigaridas, Stanimire Tomov, and Benjamin Kimia. GPU-based homotopy continuation for minimal problems in computer vision. In *Computer Vision and Pattern Recognition (CVPR)*, pages 15765–15776, 2022. 2, 3

[8] Ondřej Chum, Jiří Matas, and Josef Kittler. Locally optimized ransac. In *Pattern Recognition: 25th DAGM Symposium, Magdeburg, Germany, September 10-12, 2003. Proceedings 25*, pages 236–243. Springer, 2003. 4, 5

[9] Ondrej Chum, Tomas Werner, and Jiri Matas. Two-view geometry estimation unaffected by a dominant plane. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 772–779. IEEE, 2005. 3, 6

[10] Andrea Porfiri Dal Cin, Timothy Duff, Luca Magri, and Tomas Pajdla. Minimal perspective autocalibration. In *Computer Vision and Pattern Recognition (CVPR)*, pages 5064–5073, 2024. 2, 3

[11] David A Cox, John Little, and Donal O'shea. *Using algebraic geometry*. Springer Science & Business Media, 2005. 4, 5

[12] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018. 5, 7

[13] Yaqing Ding, Chiang-Heng Chien, Viktor Larsson, Karl Åström, and Benjamin Kimia. Minimal solutions to generalized three-view relative pose problem. In *International Conference on Computer Vision (ICCV)*, 2023. 2, 3

[14] Yaqing Ding, Jian Yang, Viktor Larsson, Carl Olsson, and Kalle Åström. Revisiting the p3p problem. In *Computer Vision and Pattern Recognition (CVPR)*, 2023. 6, 8

[15] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. Homography-based minimal-case relative pose estimation with known gravity direction. *IEEE transactions on pattern analysis and machine intelligence*, 2022. 2, 5

[16] Walter Gellert, M. Hellwich, H Kästner, and H Küstner. *The VNR concise encyclopedia of mathematics*. Springer Science & Business Media, 2012. 5

[17] Daniel R. Grayson and Michael E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www2.macaulay2.com. 4, 5

[18] Richard Hartley. Estimation of relative camera positions for uncalibrated cameras. In *European Conference on Computer Vision (ECCV)*. Springer, 1992. 1, 2

[19] Richard Hartley and Hongdong Li. An efficient hidden variable approach to minimal-case camera motion estimation. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 34(12):2303–2314, 2012. 1, 3, 5

[20] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2, 3, 5

[21] Richard I Hartley. In defense of the eight-point algorithm. *IEEE Transactions on pattern analysis and machine intelligence*, 19(6):580–593, 1997. 1

[22] Jonathan D Hauenstein and Margaret H Regan. Adaptive strategies for solving parameterized systems using homotopy continuation. *Applied Mathematics and Computation*, 2018. 2

[23] Janne Heikkilä. Using sparse elimination for solving minimal problems in computer vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 76–84, 2017. 2, 4, 6, 7, 8

[24] Petr Hruby, Timothy Duff, Anton Leykin, and Tomas Pajdla. Learning to Solve Hard Minimal Problems. In *Computer Vision and Pattern Recognition (CVPR)*, pages 5532–5542, 2022. 1

[25] Viktor Kocur, Daniel Kyselica, and Zuzana Kukelova. Robust self-calibration of focal lengths from the fundamental matrix. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5220–

5229, 2024. 3, 8

[26] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Automatic generator of minimal problem solvers. In *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part III 10*, pages 302–315. Springer, 2008. 5

[27] Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. Polynomial eigenvalue solutions to minimal problems in computer vision. *Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 2012. 1, 3, 5, 6, 7

[28] Zuzana Kukelova, Joe Kileel, Bernd Sturmfels, and Tomas Pajdla. A clever elimination strategy for efficient minimal solvers. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 1, 3, 4

[29] Viktor Larsson, Kalle Åström, and Magnus Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Computer Vision and Pattern Recognition (CVPR)*, 2017. 4

[30] Viktor Larsson and contributors. PoseLib - Minimal Solvers for Camera Pose Estimation, 2020. 6, 7, 8

[31] Viktor Larsson, Magnus Oskarsson, Kalle Åström, Alge Wallis, Zuzana Kukelova, and Tomas Pajdla. Beyond gröbner bases: Basis selection for minimal solvers. In *Computer Vision and Pattern Recognition (CVPR)*, 2018. 4, 5

[32] Hongdong Li. A simple solution to the six-point two-view focal-length problem. In *European Conference on Computer Vision (ECCV)*. Springer, 2006. 1

[33] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17627–17638, 2023. 5, 7

[34] Ezio Malis and Roberto Cipolla. Camera self-calibration from unknown planar structures enforcing the multiview constraints between collineations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002. 2

[35] Stephen Maybank. *Theory of reconstruction from image motion*. Springer, 1993. 2

[36] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 2004. 1, 2, 4, 5, 6

[37] David Nistér and Frederik Schaffalitzky. Four Points in Two or Three Calibrated Views: Theory and Practice. *International Journal of Computer Vision*, 2006. 1

[38] Long Quan, Bill Triggs, and Bernard Mourrain. Some Results on Minimal Euclidean Reconstruction from Four Points. *Journal of Mathematical Imaging and Vision*, 2006. 1

[39] Andrew J Sommese and Charles W Wampler. *The Numerical solution of systems of polynomials arising in engineering and science*. World Scientific, 2005. 2

[40] P Stefanovic. Relative orientation–a new approach. *ITC Journal*, 3(417-448):2, 1973. 4

[41] Henrik Stewénius, David Nistér, Fredrik Kahl, and Frederik Schaffalitzky. A minimal solution for relative pose with unknown focal length. In *Computer Vision and Pattern Recognition (CVPR)*, 2005. 1, 3

[42] Akihiko Torii, Zuzana Kukelova, Martin Bujnak, and Tomas Pajdla. The six point algorithm revisited. In *Computer Vision–ACCV 2010 Workshops: ACCV 2010 International Workshops, Queenstown, New Zealand, November 8-9, 2010, Revised Selected Papers, Part II 10*, pages 184–193. Springer, 2011. 3, 6, 7

[43] Charalambos Tzamos, Daniel Barath, Torsten Sattler, and Zuzana Kukelova. Relative pose of three calibrated and partially calibrated cameras from four points using virtual correspondences. *arXiv preprint arXiv:2303.16078*, 2023. 7

[44] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. 2, 7