Robust Self-calibration of Focal Lengths from the Fundamental Matrix

Viktor Kocur[1,2] Daniel Kyselica[1] Zuzana Kukelova[2]
[1] Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava
[2] Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
{viktor.kocur, daniel.kyselica}@fmph.uniba.sk kukelzuz@fel.cvut.cz

Abstract

The problem of self-calibration of two cameras from a given fundamental matrix is one of the basic problems in geometric computer vision. Under the assumption of known principal points and square pixels, the Bougnoux formula offers a means to compute the two unknown focal lengths. However, in many practical situations, the formula yields inaccurate results due to commonly occurring singularities. Moreover, the estimates are sensitive to noise in the computed fundamental matrix and to the assumed positions of the principal points. In this paper, we therefore propose an efficient and robust iterative method to estimate the focal lengths along with the principal points of the cameras given a fundamental matrix and priors for the estimated camera intrinsics. In addition, we study a computationally efficient check of models generated within RANSAC that improves the accuracy of the estimated models while reducing the total computational time. Extensive experiments on real and synthetic data show that our iterative method brings significant improvements in terms of the accuracy of the estimated focal lengths over the Bougnoux formula and other state-of-the-art methods, even when relying on inaccurate priors. The code for the methods and experiments is available at https://github.com/kocurvik/robust_self_calibration

1. Introduction

Camera calibration, *i.e.*, the estimation of camera intrinsic parameters, is a fundamental problem in computer vision with many applications. The precision of estimated intrinsic parameters, such as focal length and principal point, significantly affects the precision of tasks in structure-from-motion (SfM) [38],visual localization [34], 3D object detection [45], augmented reality [6], and other applications [39].

Classical calibration methods [46] use known calibration patterns, *e.g.* checkerboards, or additional knowledge of the observed scene to estimate the camera intrinsics. As such, they are often impractical. In contrast, self-calibration methods do not require additional knowledge of the scene geometry and rely on automatic detection of image features. Thus, they are preferable in many applications.

In this paper, we study the problem of self-calibration of two cameras, with potentially different intrinsic parameters that capture the same scene. Although this is a well-studied problem, with first solutions dated back to the nineties, to the best of our knowledge, all its existing solutions suffer from some instabilities and robustness problems.

The geometry of two uncalibrated cameras is captured by the fundamental matrix [28]. In 1998 Bougnoux [2] showed that assuming square pixels and known principal points of both cameras, it is possible to obtain the focal lengths of the two cameras from the fundamental matrix using a closed-form formula. This provides a convenient way to determine the focal length, as the principal point can usually be assumed to lie in the image center. However, the precision of the estimated focal lengths using the Bougnoux formula [2] and other similar methods [10, 18, 29] is often marred by inaccuracies, sometimes even resulting in physically implausible imaginary focal lengths. This stems from singularities and susceptibility of these methods to noise in fundamental matrices and assumed positions of principal points. Moreover, singularities occur in very common camera configurations when the principal axes of the cameras are coplanar (*i.e.* they intersect), which is common when images of a single object of interest are taken from different views.

An alternative approach [15] relies on iterative optimization of a multi-term loss, allowing one to estimate both the focal lengths of the cameras as well as their principal points. This approach addresses some of the drawbacks of the aforementioned methods but still relies on the Bougnoux formula [2] within the iterative process. Thus, it is susceptible to inaccuracies, especially in the aforementioned degenerate camera configurations.

Similarly to [15], we formulate the problem of estimating focal lengths and principal points from a given fundamental matrix as a constrained optimization problem using priors. Constraints ensure that the solution satisfies the Kruppa equations [27]. To solve the constrained optimization problem, we search for the stationary points of its La-

grangian. This results in a complex system of polynomial equations that we efficiently solve in an iterative way. Each iteration in our formulation involves solving a system of two polynomial equations of degree four in two variables, which is efficiently solved using the Gröbner basis method [24].

Existing approaches, as well as our method, are based on fundamental matrices that can be estimated using different variants of RANSAC [1, 7–9, 11, 16]. Some of the matrices produced by the 7-point algorithm [14] within RANSAC lead to imaginary focal lengths when decomposed using the Bougnoux formula [2]. We observed that such matrices rarely lead to the final RANSAC model. Based on this insight, we propose performing a degeneracy check for such matrices and rejecting them inside RANSAC, thus saving computational time on their further processing, such as scoring and local optimization.

In summary, the contributions of this paper are:

- A novel iterative method for focal length and principal point estimation from fundamental matrices. Evaluation on synthetic and large-scale real-world datasets [17, 47] shows that our method results in superior accuracy of estimated focal lengths as well as camera poses.
- A simple and computationally efficient degeneracy check for fundamental matrices leading to imaginary focal lengths within RANSAC. Performance on real-world data [17, 47] shows that performing this check and rejecting such models within various RANSAC variants [1, 7–9] and implementations [3, 16, 23] leads to faster computations and more accurate pose and focal length estimates.

2. Related Work

The problem of self-calibration of a camera from a given fundamental matrix is a well-studied problem, with several solutions that can be divided into two main groups, the direct methods and the iterative methods.

2.1. Direct Methods

Bougnoux [2] showed that under the assumption of square pixels and known positions of the principal points, it is possible to calculate the focal lengths of both cameras from the fundamental matrix in closed form. Kanatani and Matsunuga [18] derived a closed form solution directly from the elements of the fundamental matrix, avoiding the intermediate epipole computation required by the Bougnoux formula. Melanitis and Maragos [29] formulated a linear system, the solution of which provides the focal lengths.

These approaches are all algebraically equivalent and thus suffer from the same generic singularity, which occurs whenever the principal axes of the two cameras are coplanar, *i.e*. they are intersecting. This is a very common scenario in practice, as human photographers tend to place objects of interest in the centers of images. The singularity significantly affects the accuracy of the estimated focal lengths even when the axes are close to coplanar.

Similar formulas have been derived for the case when the focal lengths of the cameras are known to be equal [4, 18, 41]. In case of equal focal lengths, it is possible to avoid estimating the full 7-DoF fundamental matrix and instead employ a 6-point minimal solver [40] within RANSAC to find the relative poses of the cameras along with the focal length. These methods also suffer from generic singularities when the principal axes are parallel or intersecting and the camera centers are equidistant from the intersection point [42].

When only one of the focal lengths is unknown, the focal length can be estimated in closed form from the fundamental matrix [44] or using a 6-point minimal solver [5]. A degenerate case for these methods only occurs when the center of the calibrated camera lies on the principal axis of the uncalibrated camera, a situation rarely occurring in practice.

Since focal lengths appear in squared form in the Kruppa equations [27], all the previously mentioned direct methods provide only the squares of focal lengths. This may result in physically implausible imaginary focal lengths due to noise. Depending on the level of noise and camera configuration, imaginary focal lengths can occur at a significant rate (over 20% of samples) [19]. Imaginary focal lengths are less likely to occur when self-calibration is performed using three views [20].

2.2. Iterative Methods

Hartley and Silpa-Anan [15] pointed out that errors in the assumed positions of principal points may significantly affect the focal lengths estimated by the Bougnoux formula. To remedy this issue, they proposed an iterative method based on LM optimization [25]. During the optimization, the fundamental matrix and the positions of the principal points in both images are optimized to minimize a multi-term loss. The loss consists of the Sampson error on the correspondences and the distance of the principal points and squared focal lengths from predetermined priors. To prevent imaginary focal lengths, the loss also contains a penalization term which dramatically increases whenever the square focal lengths are below a given threshold. During the optimization procedure, the focal lengths are estimated from the fundamental matrix using the Bougnoux formula. This approach overcomes many of the issues of using the formula alone, but may still suffer in the vicinity of degenerate camera configurations, as it still relies on the formula to estimate the focal lengths. Additionally, the procedure is computationally expensive as it requires computation of the Sampson error for all correspondences in each iteration.

Iterative methods can also be used to estimate the camera intrinsics using two or more views [10, 12, 27, 31]. When considering only two views with two unknown focal lengths, these methods should converge to the Bougnoux formula. However, these methods may result in different

outcomes in the vicinity of degenerate configurations. For example, the method by Fetzer *et al*. [10] optimizes for an energy functional based on the Kruppa equations [27]. In degenerate configurations, the functional does not have a single minimum, yet the optimization procedure will be terminated at some point, providing an output different from the one provided by the Bougnoux formula.

Next, we describe the proposed iterative method that aims to avoid the aforementioned problems of existing approaches for self-calibration from the fundamental matrix.

3. Robust focal length estimation

In this paper, similar to Hartley [15], we formulate the problem of estimating focal lengths from a given fundamental matrix $\mathbf{F}$ as an optimization problem using priors.

Let $\mathbf{F} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$ be the SVD of the fundamental matrix $\mathbf{F}$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ are orthonormal matrices and $\mathbf{D} = \text{diag}(\sigma_1, \sigma_2, 0)$ is a diagonal matrix with two non-zero singular values σ_1 and σ_2 of $\mathbf{F}$. Let $\omega_i^* = \mathbf{K}_i\mathbf{K}_i^\top$ be the dual image of the absolute conic for the i^{th} camera, $i = 1, 2$, with $\mathbf{K}_i$ being the 3×3 calibration matrix of the form

$$\mathbf{K}_i = \begin{bmatrix} f_i & 0 & u_i \\ 0 & f_i & v_i \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

with the focal length f_i and principal point $\mathbf{c}_i = [u_i, v_i]^\top$.

For fixed principal points, *e.g.*, $\mathbf{c}_i = [0, 0]$, $i = 1, 2$, there is a closed-form solution for the focal lengths extracted from the given fundamental matrix $\mathbf{F}$, *i.e.*, the Bougnoux formula [2]. Since this formula can result in unstable or imaginary estimates, in our formulation, we allow moving the estimated principal points to avoid such instabilities. For non-fixed principal points $\mathbf{c}_i$, there are infinitely many decompositions of $\mathbf{F}$ into the essential matrix $\mathbf{E}$ and two calibration matrices of the form (1). Thus, we formulate the problem of estimating focal lengths from $\mathbf{F}$ as a constrained optimization, where during the optimization, the focal lengths, along with the positions of principal points in both images are optimized to minimize the cost function:

$$\begin{aligned} \min_{f_1, f_2, \mathbf{c}_1, \mathbf{c}_2} \quad & \sum_{i=1,2} w_i^f (f_i - f_i^p)^2 + w_i^\mathbf{c} \|\mathbf{c}_i - \mathbf{c}_i^p\|^2 \\ \text{s.t.} \quad \kappa_1 = \; & \sigma_1(\mathbf{v}_1^\top \omega_1^* \mathbf{v}_1)(\mathbf{u}_1^\top \omega_2^* \mathbf{u}_2) + \\ & + \sigma_2(\mathbf{v}_1^\top \omega_1^* \mathbf{v}_2)(\mathbf{u}_2^\top \omega_2^* \mathbf{u}_2) = 0 \\ \kappa_2 = \; & \sigma_1(\mathbf{v}_1^\top \omega_1^* \mathbf{v}_2)(\mathbf{u}_1^\top \omega_2^* \mathbf{u}_1) + \\ & + \sigma_2(\mathbf{v}_2^\top \omega_1^* \mathbf{v}_2)(\mathbf{u}_1^\top \omega_2^* \mathbf{u}_2) = 0, \end{aligned} \tag{2}$$

where f_i^p and $\mathbf{c}_i^p = [u_i^p, v_i^p]^\top$, $i = 1, 2$ are the priors for the focal lengths and the principal points, w_i^f and $w_i^\mathbf{c}$ are predetermined weights, and $\kappa_1 = 0$ and $\kappa_2 = 0$ are two Kruppa equations [27][1], which are functions of both focal lengths and principal points that appear in $\omega_i^* = \mathbf{K}_i\mathbf{K}_i^\top$. The Kruppa equations ensure that for the input fundamental matrix $\mathbf{F}$ and the estimated calibration matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ (1), the matrix $\mathbf{K}_1^\top \mathbf{F} \mathbf{K}_2$ is a valid essential matrix.

The cost function in (2) has several advantages over the cost function used in [15]. (1) It does not contain a term with the Sampson error on correspondences, which significantly slows down the optimization in [15]. (2) It operates directly on focal lengths f_i instead of squared focal lengths f_i^2. (3) The minimization of (2) does not require computation or initialization using the Bougnoux formula [2] as used in [15]. Such an initialization can be very far from the ground truth focal lengths and in many scenarios can result in $f_i^2 \leq 0$. Therefore, minimization of (2) does not need a special penalty term for $f_i^2 \leq 0$ as used in [15].

The constrained optimization problem (2) can be transformed into an unconstrained problem using the method of Lagrange multipliers. This leads to the Lagrange function $L(f_1, f_2, \mathbf{c}_1, \mathbf{c}_2, \lambda_1, \lambda_2)$:

$$L = \sum_{i=1,2} w_i^f (f_i - f_i^p)^2 + w_i^\mathbf{c} \|\mathbf{c}_i - \mathbf{c}_i^p\|^2 - 2\lambda_1\kappa_1 - 2\lambda_2\kappa_2, \tag{3}$$

where λ_1 and λ_2 are the Lagrange multipliers. The constant '-2' is introduced for an easier subsequent manipulation of the equations and does not influence the final solution.

In this case, if $(f_1^\star, f_2^\star, \mathbf{c}_1^\star, \mathbf{c}_2^\star)$ is a point of the minimum of the original constrained problem (2), then there exist $\lambda_1^\star$ and $\lambda_2^\star$ such that $(f_1^\star, f_2^\star, \mathbf{c}_1^\star, \mathbf{c}_2^\star, \lambda_1^\star, \lambda_2^\star)$ is a stationary point of L (3), *i.e.*, a point where all the partial derivatives of L vanish. The Lagrange function $L(f_1, f_2, \mathbf{c}_1, \mathbf{c}_2, \lambda_1, \lambda_2)$ (3) is a function of eight unknowns. Thus, to find all stationary points of L we need to solve the following system of eight polynomial equations in eight unknowns:[2]

$$2w_i^f(f_i - f_i^p) - 2\lambda_1 \frac{\partial \kappa_1}{\partial f_i} - 2\lambda_2 \frac{\partial \kappa_2}{\partial f_i} = 0, \tag{4}$$

$$2w_i^\mathbf{c}(\mathbf{c}_i - \mathbf{c}_i^p) - 2\lambda_1 \left(\frac{\partial \kappa_1}{\partial \mathbf{c}_i}\right)^\top - 2\lambda_2 \left(\frac{\partial \kappa_2}{\partial \mathbf{c}_i}\right)^\top = \mathbf{0}, \tag{5}$$

$$\kappa_1 = 0, \tag{6}$$

$$\kappa_2 = 0. \tag{7}$$

Unfortunately, the system of eight equations (4)-(7) is too complex to be efficiently solved using an algebraic method, *e.g.* using the Gröbner basis method [22, 24]. Next, we propose an iterative method to efficiently find a solution of system (4)-(7) that minimizes the cost function (2).

[1] From three Kruppa equations only two are linearly independent.

[2] Note that (4) represents two equations for $i = 1, 2$ and (5) represents four equations, since there are two vector equations for $i = 1, 2$.

3.1. Iterative method

First, let us denote $\Delta f_i = f_i - f_i^p$, $\Delta \mathbf{c}_i = \mathbf{c}_i - \mathbf{c}_i^p$, $i = 1, 2$. From equations (4)–(5) we have

$$\Delta f_i = \frac{1}{w_i^f}(\lambda_1 \frac{\partial \kappa_1}{\partial f_i} + \lambda_2 \frac{\partial \kappa_2}{\partial f_i}), \tag{8}$$

$$\Delta \mathbf{c}_i = \frac{1}{w_i^{\mathbf{c}}}\left(\lambda_1 \left(\frac{\partial \kappa_1}{\partial \mathbf{c}_i}\right)^\top + \lambda_2 \left(\frac{\partial \kappa_2}{\partial \mathbf{c}_i}\right)^\top\right). \tag{9}$$

and the cost function in the original constrained optimization problem (2) can be rewritten as

$$e = \sum_{i=1,2} w_i^f \Delta f_i^2 + w_i^{\mathbf{c}} \Delta \mathbf{c}_i^\top \Delta \mathbf{c}_i. \tag{10}$$

The proposed iterative solution to equations (4)–(7) follows the idea of iterative triangulation methods [21, 26]. First, let $s^{k-1} = \langle f_1^{k-1}, f_2^{k-1}, \mathbf{c}_1^{k-1}, \mathbf{c}_2^{k-1} \rangle$ denote the best current estimate of focal lengths and principal points after the $(k-1)^{th}$ iteration. The prior values $f_i^0 \equiv f_i^p$, $\mathbf{c}_i^0 \equiv \mathbf{c}_i^p$, $i = 1, 2$ are used as an initialization. In the k^{th} iteration the updated estimates $s^k = \langle f_1^k, f_2^k, \mathbf{c}_1^k, \mathbf{c}_2^k \rangle$ are obtained by replacing the focal lengths and principal points $\langle f_1, f_2, \mathbf{c}_1, \mathbf{c}_2 \rangle$ on the right-hand side of equations (8)–(9) by the current best estimates $s^{k-1} = \langle f_1^{k-1}, f_2^{k-1}, \mathbf{c}_1^{k-1}, \mathbf{c}_2^{k-1} \rangle$. In this way, we obtain the expressions for $\Delta f_i = \frac{1}{w_i^f}(\lambda_1 \frac{\partial \kappa_1}{\partial f_i}(s^{k-1}) + \lambda_2 \frac{\partial \kappa_2}{\partial f_i}(s^{k-1}))$ and $\Delta \mathbf{c}_i = \frac{1}{w_i^{\mathbf{c}}}(\lambda_1 (\frac{\partial \kappa_1}{\partial \mathbf{c}_i}(s^{k-1}))^\top + \lambda_2 (\frac{\partial \kappa_2}{\partial \mathbf{c}_i}(s^{k-1}))^\top)$, $i = 1, 2$. Here $\frac{\partial \kappa_j}{\partial f_i}(s^{k-1})$ and $\frac{\partial \kappa_j}{\partial \mathbf{c}_i}(s^{k-1})$, $i, j = 1, 2$ are partial derivatives of κ_j evaluated at $s^{k-1} = \langle f_1^{k-1}, f_2^{k-1}, \mathbf{c}_1^{k-1}, \mathbf{c}_2^{k-1} \rangle$. After this substitution, Δf_i and $\Delta \mathbf{c}_i$ are functions of two Lagrange multipliers λ_1 and λ_2.

Expressions Δf_i and $\Delta \mathbf{c}_i$ are, in turn, substituted into the Kruppa equations (6) and (7). This is done by substituting Δf_i and $\Delta \mathbf{c}_i$, $i = 1, 2$ into ω_i^*using the relationship $f_i = \Delta f_i + f_i^p$ and $\mathbf{c}_i = \Delta \mathbf{c}_i + \mathbf{c}_i^p$.

The updated Kruppa equations (6) and (7), which we denote $\kappa_1^k = 0$ and $\kappa_2^k = 0$, form a system of two equations of degree four in two unknowns λ_1 and λ_2. This system has up to 16 real solutions and can be efficiently solved using the Gröbner basis method [24]. The final solver performs an elimination of a 20×36 matrix and extracts solutions from the eigenvalues and eigenvectors of a 16×16 matrix.

From up to 16 possible solutions, a solution λ_1^k and λ_2^k that minimizes $|\lambda_1| + |\lambda_2|$ is selected [26]. This solution is used to compute the k^{th} iteration updates Δf_i^k and $\Delta \mathbf{c}_i^k$ using equations (8)–(9) and subsequently the new estimates of focal lengths and principal points $s^k = \langle f_1^k, f_2^k, \mathbf{c}_1^k, \mathbf{c}_2^k \rangle$.

The iterative algorithm stops when the relative change of error of two consecutive iterations $\frac{|e^k - e^{k-1}|}{e^k} < \epsilon$, for a given threshold ϵ. Note that in each iteration, the obtained solutions satisfy the Kruppa equations $\kappa_1 = 0$ and $\kappa_2 = 0$. This means that for the input fundamental matrix $\mathbf{F}$ and the estimated calibration matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ (1), the matrix $\mathbf{K}_1^\top \mathbf{F} \mathbf{K}_2$ is a valid essential matrix in each iteration.

The situation with $f_1 = f_2$ can be solved using the same method. In this case, in the cost function (2), as well as in the Lagrangian and in (4)–(5), we only have $i = 1$. However, since we still have two Kruppa equations $\kappa_1 = 0$ and $\kappa_2 = 0$ the final algorithm also relies on solutions to a system of two equations of degree four in two unknowns. The algorithm is summarized in the Supplementary material.

3.2. Real Focal Length Checking within RANSAC

The method proposed in 3.1 is usually applied on a fundamental matrix $\mathbf{F}$ that is obtained using a robust RANSAC-style estimation [1, 7–9, 11, 16]. Compared to the Bougnoux formula [2], our method does not suffer from imaginary focal length estimates in the presence of noise or an error in the principal point location. Thus, our method does not need to check whether the fundamental matrix returned by the RANSAC is decomposable to real focal lengths or not. Nevertheless, as we observed in our experiments, fundamental matrices that result in imaginary focal lengths when decompossed with the Bougnoux formula usually do not provide good estimates and are therefore not selected inside RANSAC as best models.

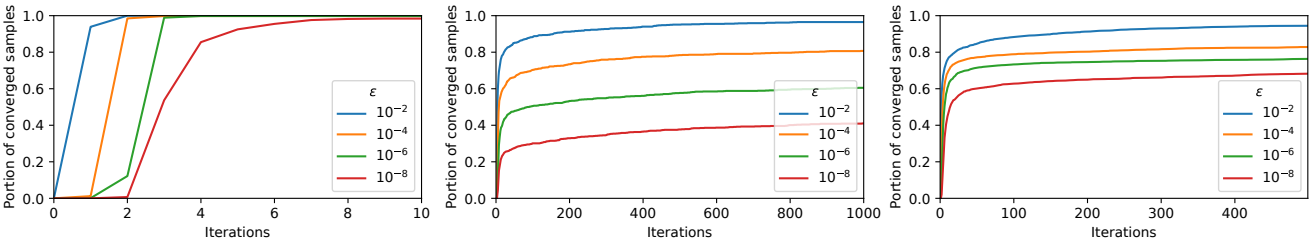
Inspired by this observation, we propose a simple modification of RANSAC that is based on rejecting models that lead to imaginary focal lengths. Rejection is done before scoring models on all point correspondences. This reduces the computational time required to score a model, which is unlikely to lead to the final output. Thus, this approach is similar to other degeneracy checking algorithms.

To reject models, we use a modified version of the Bougnoux formula [2] presented in [32], which has the form of a ratio of two polynomials in the elements of the fundamental matrix. The formula returns squared focal lengths. When the sign of one of the squared focal lengths is negative, we reject the model. This approach does not require the use of SVD or other expensive matrix manipulations. This type of checking can be implemented into different variants of RANSAC and, as shown later in our experiments, it leads to reduced computational times while maintaining or even improving accuracy of pose estimates. Implementation details can be found in the Supplementary Material (SM).

4. Synthetic Experiments

We perform several synthetic experiments to evaluate various aspects of our method and compare it with the state-of-the-art in controlled configurations. Here, we only include experiments for two unknown but different focal lengths. Further experiments for the case where the focal lengths are assumed to be equal can be found in the SM.

In all our synthetic and real experiments, we set the weights in (2) to $w_i^f = 5 \cdot 10^{-4}$ and $w_i^{\mathbf{c}} = 1.0$. These



(a) $\mathcal{C}(0°, 300)$ (b) $\mathcal{C}(0°, 0)$ (c) Phototourism [17]

Figure 1. Plots showing portion of samples for which our algorithm would converge given a threshold for the relative change of errors in successive iterations $\frac{|e^n - e^{n-1}|}{e^n} < \epsilon$. For synthetic experiments (a) and (b) we generated 1000 samples with added noise ($\sigma_n = 1$, $\sigma_p = 10$). We set the priors as $f_1^p = 660$, $f_2^p = 440$. For (c) we used the Phototourism benchmark dataset [17]. See Section 5 for details.

weights were obtained by testing different combinations on the Brandenburg Gate validation scene from the Phototurism dataset [17] and resulted in the best performance on this scene. However, as we show also in synthetic experiments, for a wide variety of combinations of weights, we consistently obtain very good estimates and outperform the state-of-the-art methods. We set the weights for Hartley and Silpa-Anan's method [15] in the same manner to 1.0 for w_F, 10^{-4} for $w_{\mathbf{c}}$ and 10^{-6} for w_f.

Camera Setup: To perform the experiments, we set up a scene with two cameras. We assume 640×480 image size and principal points in image centers. The first camera has focal length $f_1 = 600$ and the second $f_2 = 400$. The first camera has center at the origin with the z-axis of the world coordinate system being its principal axis. We perform experiments with the second camera in different configurations denoted as $\mathcal{C}(\theta, y)$. For $\mathcal{C}(0°, y)$ the second camera center is located at coordinates $(1200, y, 600)$. The camera is rotated by $60°$ around its y-axis. With $\theta \neq 0°$ we additionally rotate the second camera around the x-axis by θ. When $y = 0$ and $\theta = 0°$ the principal axes of the two cameras intersect and are thus coplanar, which is a degenerate configuration for the Bougnoux formula [2].

For a given camera setup, we uniformly generate 100 random points visible by both cameras. After projecting the points into both images, we add Gaussian noise with a standard deviation σ_n to the pixel coordinates. We also emulate errors in the assumed locations of the principal points by shifting them c_{err} pixels in random direction from the ground truth. The shift c_{err} is either a fixed value or sampled from a normal distribution with a standard deviation of σ_c. We estimate the fundamental matrix using MAGSAC++ [1] as implemented in OpenCV [3].

Convergence: In the first experiment, we evaluate convergence of the proposed iterative method. Different termination criteria can be considered. We terminate the algorithm when the relative change in error in two successive iterations $\frac{|e^n - e^{n-1}|}{e^n}$ is less than a predetermined threshold. Fig. 1a shows that for non-degenerate configuration our method converges quickly even with strict thresholds. Fig. 1b shows the convergence rates of our method when the principal axes are coplanar. The degenerate configuration requires an increased number of iterations to converge with the same thresholds. However, even in a degenerate configuration, the method converges in a few iterations for most of the samples, even with strict thresholds. We also evaluate convergence on a large scale real-world dataset [17] (see section 5 for more details). As shown in Fig. 1c, our method successfully converges for most of the samples even with a strict threshold. Using these findings, we established the maximum number of iterations for our method at 50, as further iterations demonstrated diminishing returns in terms of convergence rate.

Accuracy of the Estimated Focal Lengths In the next experiment, we compare our method with state-of-the-art methods in terms of accuracy of the estimated focal lengths. The methods considered are the baseline Bougnoux formula [2] and the iterative approaches of Hartley and Silpa-Anan [15] and Fetzer *et al*. [10]. In the experiments, we use the same priors for Hartley and Silpa-Anan's method as for ours. We set the initial values for LM optimization for the method by Fetzer *et al*. to the same values as our priors.

Fig. 2a and 2b show how the transition to a degenerate configuration with coplanar principal axes affects the accuracy. Our method performs well even close to and in degenerate configurations. In comparison, the other methods fail to provide consistent results when the camera position approaches the degenerate configuration.

Fig. 2c shows that an increase in c_{err} results in a worsening performance of the Bougnoux formula and the method by Fetzer *et al*. Our method is capable of compensating for the error in the principal point as it is not assumed to be known exactly. The performance of Hartley and Silpa-Anan's method stays consistent for similar reasons, although it performs worse than ours.

In Fig. 2d, we show the performance of the methods under varying levels of added image noise. Our method demonstrates higher robustness to noise, resulting in re-

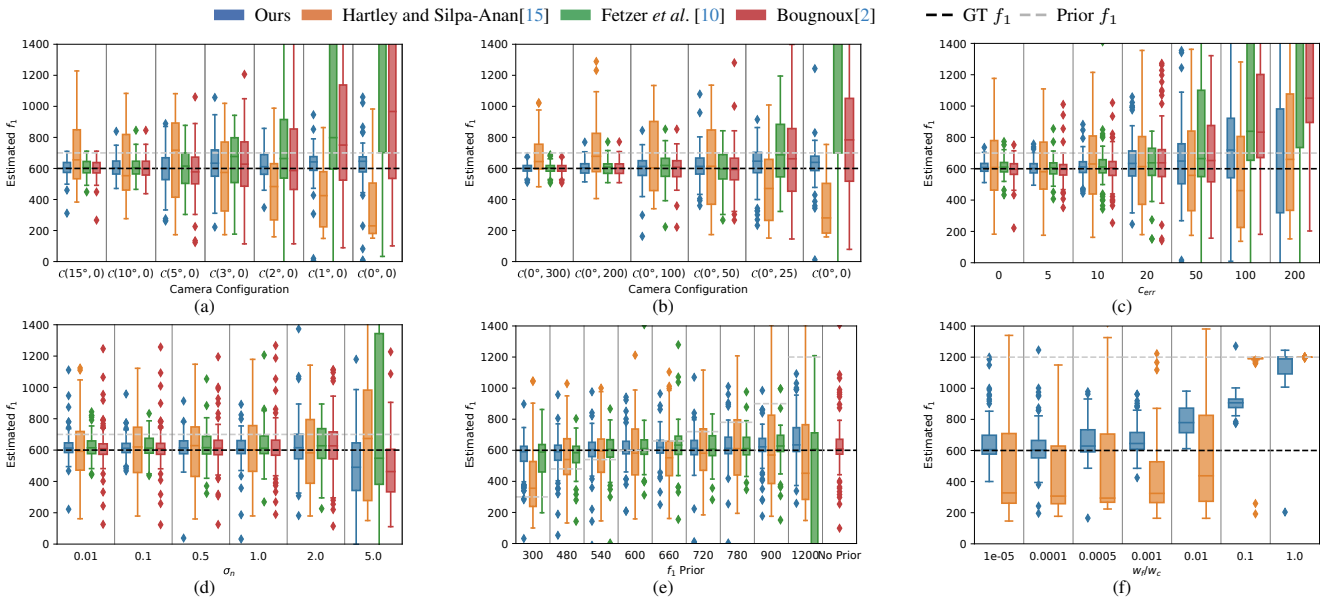


Figure 2. Synthetic experiments: Box plots for the estimated focal lengths of the first camera. Comparison of the methods as (a, b) the camera configuration approaches the degeneracy with coplanar principal axes, (c) we vary the error in principal point c_{err}, (d) we vary the noise added to the projected points, (e) we vary the prior for f_1, (f) we vary the relative weights of focal length and principal point priors. We use priors $f_1^p = 700$, $f_2^p = f_2 = 400$ for (a, b, c, d), $f_1^p = 1200$, $f_2^p = 400$ for (f), $\sigma_n = 1$ for (a, b, c, e, f), $\sigma_c = 10$ for (a, b, d, e, f). For (c, d, e, f) we randomly sample the configuration $\mathcal{C}(\theta, y)$ with $\theta \in [-15°, 15°]$ and $y \in [-200, 200]$.

duced inter-sample deviation and a more consistent median estimate across the various noise levels. Fig. 2e shows how the different methods respond to the priors for focal length, or in the case of the method by Fetzer *et al*. to an initialization for the iterative process. Fig. 2f shows how the setting of weights for our and Hartley and Silpa-Anan's method affects the accuracy of the estimated focal lengths.

5. Real-world Experiments

Datasets: We assess the real-world performance of our method using two extensive datasets. The first dataset, referred to as the **Phototourism dataset**, was originally introduced in [17] to serve as a robust benchmark to evaluate structure-from-motion pipelines. The dataset contains 25 scenes of landmarks. Large sets of crowd-sourced images are available for each scene. For 13[3] of the scenes a COLMAP [36] reconstruction is provided to serve as ground truth for camera intrinsics and extrinsics. To evaluate the estimated focal lengths, we randomly sample 1000 pairs of images for each scene. We only consider pairs with sufficient co-visibility as defined in [17]. We used the Brandenburg Gate scene to serve as a validation set to determine the optimal setting of the weights w_i^f and w_i^c in (2).

The second dataset, known as the **Aachen Day-Night v1.1 Dataset**, is an extension of the previously established

[3]Originally, reconstructions were available for 16 of the scenes, but 3 of the scenes were later dropped from the dataset due to data inconsistencies.

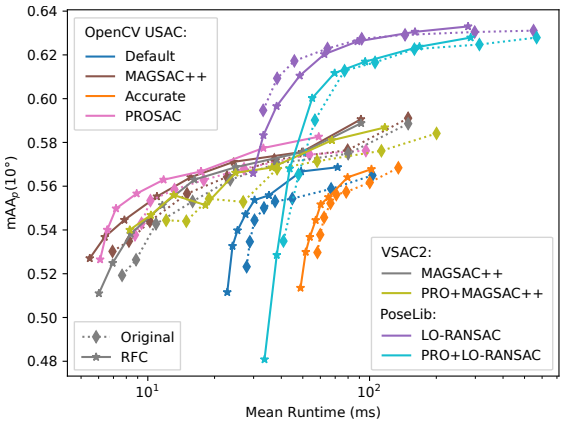
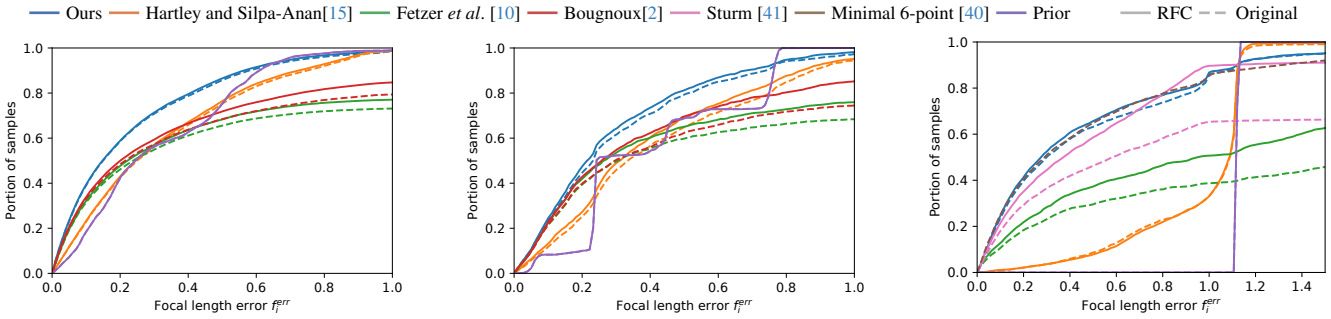


Figure 3. The plot shows mAA$_p$ achieved on the Phototourism dataset [17] for different RANSAC implementations [3, 16, 23] with and without real focal length checking (RFC) under varying number of iterations. To calculate the poses we use the ground truth focal lengths. Further details are provided in the SM.

Aachen Day-Night dataset [33, 35]. This dataset incorporates a reference COLMAP [36] model that has been reconstructed from a comprehensive collection of 6,697 images captured by a variety of cameras and smartphones. For evaluation, we randomly sample 1000 pairs of images that adhere to the same co-visibility criteria as those defined for the Phototourism dataset. COLMAP ground truth can introduce some bias; however, it is obtained using optimization



(a) Phototourism [17] (b) Aachen Day-Night v1.1 [47] (c) ETH3D Multiview [37]

Figure 4. Plots showing the portion of samples for which the estimated focal lengths were below a given f_i^{err} threshold. For (a) and (b) we assume different focal lengths for the two cameras, for (c) the focal lengths are assumed to be equal.

Method	RFC	Phototourism [17]						Aachen Day-Night v1.1 [47]					
		Median p_{err}	mAA$_p$ 10°	mAA$_p$ 20°	Median f_{err}	mAA$_f$ 0.1	mAA$_f$ 0.2	Median p_{err}	mAA$_p$ 10°	mAA$_p$ 20°	Median f_{err}	mAA$_f$ 0.1	mAA$_f$ 0.2
Ours		6.46°	40.37	55.90	0.149	23.95	37.29	9.53°	27.55	46.17	0.232	12.46	23.87
	✓	**6.31°**	**40.97**	**56.70**	**0.148**	**24.23**	**37.57**	**8.78°**	**29.30**	**48.11**	**0.214**	**12.73**	**24.97**
Hartley [15]		9.19°	30.15	46.94	0.250	12.73	23.39	12.19°	21.10	38.74	0.336	6.37	12.68
	✓	9.00°	30.61	47.69	0.244	12.88	23.67	11.37°	22.12	40.61	0.312	7.19	14.19
Fetzer [10]		9.40°	33.00	47.25	0.238	19.53	29.89	12.33°	24.34	39.69	0.305	11.47	21.50
	✓	8.94°	33.64	48.51	0.221	19.93	30.66	10.66°	25.62	42.16	0.263	11.51	22.63
Bougnoux [2]		7.55°	37.17	52.70	0.215	20.65	31.47	10.25°	26.73	45.07	0.295	11.37	21.64
	✓	7.39°	37.57	53.21	0.200	21.05	32.26	9.69°	27.25	45.61	0.259	11.78	22.96
Prior		11.17°	22.63	41.98	0.246	8.83	19.81	13.00°	17.06	37.27	0.242	4.11	6.55
	✓	11.05°	22.73	42.31	0.246	8.83	19.81	12.73°	17.68	38.05	0.242	4.11	6.55
GT intrinsics		2.85°	59.78	70.80	—	—	—	5.45°	45.99	58.66	—	—	—
	✓	2.82°	60.05	71.18				4.25°	51.76	64.82			

Table 1. Median errors for poses (p_{err}) and focal lengths (f_i^{err}) and mean average accuracy scores for poses (mAA$_p$) and estimated focal lengths (mAA$_f$) on 12 scenes from the Phototourism dataset [17] and the Aachen Day-Night v1.1 dataset [47]. RFC denotes the real focal length checking as described in subsection 3.2.

(BA) over many images. This is widely accepted for generating ground truth and is expected to be more precise than the estimates obtained from pairs of images.

Metrics: To evaluate the accuracy of the output focal lengths, we define relative focal length error as

$$f_i^{err} = \frac{|f_i^{est} - f_i^{gt}|}{f_i^{gt}}, \tag{11}$$

with f_i^{est} and f_i^{gt} denoting the estimated and ground truth focal lengths respectively, with $i \in \{1, 2\}$ denoting the first or the second camera. To gauge the accuracy of the poses, we employ the Mean Average Accuracy (mAA) metric, as originally proposed in [17]. The basis for this metric is the pose error (p_{err}) which is calculated as the maximum of the rotation and translation error in degrees.

Furthermore, we employ the mAA metric to evaluate and compare the accuracy of the estimated focal lengths, based on the f_i^{err} curves as presented in Figure 4. To distinguish between the two applications of the mAA metric, we denote them as mAA$_p$ for pose and mAA$_f$ for focal lengths.

Fundamental Matrix Estimation: We use LoFTR [43] to obtain point correspondences. We perform inference on images resized so that the largest dimension is equal to 1024 px. To estimate the fundamental matrix, we use MAGSAC++ [1] with the epipolar threshold set to 3 px.

We also perform experiments with and without utilizing real focal length checking (RFC) as described in subsection 3.2. Fig 3 shows that on the Phototourism dataset [17] the fundamental matrices obtained with RFC generally lead to more accurate pose estimates, while also reducing computational times across various RANSAC implementations. More details and further experiments are provided in SM.

Compared Methods: We compare our method with three state-of-the-art approaches. We evaluate the Bougnoux formula [2], Hartley and Silpa-Anan's iterative method [15] and the iterative method by Fetzer *et al*. [10].

For all images, we consider the principal point to lie in their center which is a commonly used assumption. With the exception of the Bougnoux formula, all methods require priors, or initial estimates, for the focal lengths. Following common practice, as established in COLMAP [36], we ini-

Method	Library	RFC	Mean Time (ms)	Mean iterations
Ours	Matlab		22.42	24.33
		✓	14.95	22.53
	Eigen (C++)		3.59	23.63
		✓	3.44	22.49
Hartley [15]	PyTorch [30]		659.61	85.38
		✓	663.98	86.14
Fetzer [10]	PyTorch [30]		197.89	0.00
		✓	191.75	40.56
Bougnoux [2]	NumPy [13]		0.27	—
		✓	0.25	—

Table 2. Mean computation times and number of iterations for the compared methods on the Phototourism dataset [17]. Measurements were performed on Intel i7-11800H. The times shown do not include the estimation of the fundamental matrices.

Method	Median	mAA_p		Median	mAA_f	
	p_{err}	10°	20°	f_{err}	0.1	0.2
Ours	**16.51°**	16.70	32.24	0.299	12.96	23.10
Hartley [15]	21.52°	12.38	24.26	1.090	0.45	0.99
Fetzer [10]	70.40°	2.31	5.20	0.999	6.86	11.64
Sturm [41]	17.98°	**20.64**	**33.30**	0.582	10.39	17.55
Minimal [40]	36.18°	16.42	25.74	**0.291**	**14.43**	**24.29**
Prior	24.60°	11.24	20.92	1.119	0.00	0.00
GT intrinsics	6.40°	41.73	54.16	—	—	—

Table 3. Mean average accuracy scores for poses (mAA_p) and estimated focal lengths (mAA_f) on 12 scenes from the ETH3D High Resolution Multiview dataset [37]. The focal lengths of both cameras are assumed to be equal. The last row includes reference results for pose accuracy when ground truth intrinsics are used.

tialize the priors for the focal lengths at 1.2 times the maximum dimension of the image This initial setting implies an approximate 50-degree field-of-view for the cameras.

Results: The median errors and mean average accuracy scores for both pose and focal lengths are reported in Table 1. On both datasets, our method significantly outperforms others in terms of the accuracy of the estimated focal lengths. The superiority of our method in terms of focal length accuracy can also be seen in Fig. 4a and 4b. In terms of pose accuracy, our method also provides more accurate results than all the compared methods.

The results also show that performing RFC for real focal lengths within RANSAC as proposed in subsection 3.2 generally leads to improvements in terms of both the pose and focal length accuracy across the compared methods.

Computational Speed: Table 2 shows the mean computational times for the various methods. Our method is significantly faster than other iterative methods. Note that the implementation of our method is not optimized and there is still some space for speed-up.

Equal Focal Lengths: We also evaluate our method for the case when both focal lengths can be assumed to be equal. In this case, we use the modified version of our iterative algorithm from Section 3.1. To perform the evaluation, we use the ETH3D Multiview dataset [37]. We use all image pairs satisfying the co-visibility criterion for each of the 12 scenes from the test set, yielding 3830 total pairs. The dataset combines laser scans and photometric errors to obtain the ground truth. For comparison, we modified the method by Fetzer *et al*. to optimize for a single focal length. We use Hartley and Silpa-Anan's method as usual, but on output we average the two focal lengths produced. Instead of the Bougnoux formula, we use the formula proposed by Sturm [41] which is specific to the case of equal focal lengths. We also evaluate the minimal 6-point algorithm [40] implemented within LO-RANSAC [8, 23]. Further details and additional experiments are available in SM.

The results are shown in Table 3 and Fig. 4c. Our method provides superior results to other methods based on the decomposition of the fundamental matrix, and performs comparably to the minimal solver [40] in terms of estimated focal lengths and shows minor improvements in terms of pose accuracy. Our method thus provides a good alternative to the 6-point minimal solver in the case when the focal lengths of the two cameras are known to be equal.

6. Conclusion

We address the important problem of robust self-calibration of the focal lengths of two cameras from a given fundamental matrix. Our new efficient iterative method shows substantial improvements over existing techniques. Synthetic experiments show that our method performs reliably even when the cameras are in or close to degenerate configurations. In real-world evaluations on two large-scale datasets, our approach demonstrates superior accuracy in terms of estimated poses and focal lengths, even when we use inaccurate initial priors, while also being faster than competing iterative approaches.

We have additionally proposed to perform a computationally simple check within the standard 7-point algorithm to remove models that lead to imaginary focal lengths. Real-world experiments show that this approach not only decreases computational time of the whole RANSAC pipeline, but, in general, also leads to improved pose and focal length accuracy across multiple RANSAC variants and implementations.

Acknowledgments: V. K. was supported by the Slovak Grant Agency for Science (VEGA), project no. 1/0373/23. and the TERAIS project, a Horizon-Widera-2021 program of the European Union under the Grant agreement number 101079338. 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. Z. K. was supported by the Czech Science Foundation (GAČR) JUNIOR STAR Grant No. 22-23183M.

References

[1] Daniel Barath, Jana Noskova, Maksym Ivashechkin, and Jiri Matas. Magsac++, a fast, reliable and accurate robust estimator. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1304–1312, 2020. 2, 4, 5, 7

[2] 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. 1, 2, 3, 4, 5, 6, 7, 8

[3] Gary Bradski. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11):120–123, 2000. 2, 5, 6

[4] Michael J Brooks, Lourdes de Agapito, Du Q Huynh, and Luis Baumela. Towards robust metric reconstruction via a dynamic uncalibrated stereo head. *Image and Vision Computing*, 16(14):989–1002, 1998. 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. 2

[6] Robert Castle, Georg Klein, and David W. Murray. Video-rate localization in multiple maps for wearable augmented reality. In *ISWC*, 2008. 1

[7] Ondrej Chum and Jiri Matas. Matching with prosac-progressive sample consensus. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, pages 220–226. IEEE, 2005. 2, 4

[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. 8

[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)*, pages 772–779. IEEE, 2005. 2, 4

[10] Torben Fetzer, Gerd Reis, and Didier Stricker. Stable intrinsic auto-calibration from fundamental matrices of devices with uncorrelated camera parameters. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 221–230, 2020. 1, 2, 3, 5, 6, 7, 8

[11] 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. 2, 4

[12] Riccardo Gherardi and Andrea Fusiello. Practical autocalibration. In *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part I 11*, pages 790–801. Springer, 2010. 2

[13] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. 8

[14] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge, 2nd edition, 2003. 2

[15] Richard Hartley, Chanop Silpa-Anan, et al. Reconstruction from two views using approximate calibration. In *Proc. 5th Asian Conf. Comput. Vision*, pages 338–343, 2002. 1, 2, 3, 5, 6, 7, 8

[16] Maksym Ivashechkin, Daniel Barath, and Jiří Matas. Vsac: Efficient and accurate estimator for h and f. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 15243–15252, 2021. 2, 4, 6

[17] Yuhe Jin, Dmytro Mishkin, Anastasiia Mishchuk, Jiri Matas, Pascal Fua, Kwang Moo Yi, and Eduard Trulls. Image matching across wide baselines: From paper to practice. *International Journal of Computer Vision*, 129(2):517–547, 2021. 2, 5, 6, 7, 8

[18] Kenichi Kanatani and Chikara Matsunaga. Closed-form expression for focal lengths from the fundamental matrix. In *Proc. 4th Asian Conf. Comput. Vision*, pages 128–133. Citeseer, 2000. 1, 2

[19] Kenichi Kanatani, Atsutada Nakatsuji, and Yasuyuki Sugaya. Stabilizing the focal length computation for 3-d reconstruction from two uncalibrated views. *International Journal of Computer Vision*, 66:109–122, 2006. 2

[20] Yasushi Kanazawa, Yasuyuki Sugaya, and Kenichi Kanatani. Initializing 3-d reconstruction from three views using three fundamental matrices. In *Image and Video Technology–PSIVT 2013 Workshops: GCCV 2013, GPID 2013, PAES-NPR 2013, and QACIVA 2013, Guanajuato, Mexico, October 28-29, 2013, Revised Selected Papers 6*, pages 169–180. Springer, 2014. 2

[21] Zuzana Kukelova and Viktor Larsson. Radial distortion triangulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9681–9689, 2019. 4

[22] 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. 3

[23] Viktor Larsson and contributors. PoseLib - Minimal Solvers for Camera Pose Estimation, 2020. 2, 6, 8

[24] Viktor Larsson, Kalle Astrom, and Magnus Oskarsson. Efficient solvers for minimal problems by syzygy-based reduction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 820–829, 2017. 2, 3, 4

[25] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944. 2

[26] Peter Lindstrom. Triangulation made easy. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1554–1561. IEEE, 2010. 4

[27] Manolis IA Lourakis and Rachid Deriche. *Camera self-calibration using the Kruppa equations and the SVD of the fundamental matrix: The case of varying intrinsic parameters*. PhD thesis, INRIA, 2000. 1, 2, 3

[28] Quan-Tuan Luong and Olivier D Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International journal of computer vision*, 17(1):43–75, 1996. 1

[29] Nikos Melanitis and Petros Maragos. A linear method for camera pair self-calibration. *Computer Vision and Image Understanding*, 210:103223, 2021. 1, 2

[30] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 8

[31] Marc Pollefeys, Reinhard Koch, and Luc Van Gool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International journal of computer vision*, 32(1):7–25, 1999. 2

[32] Oleh Rybkin. Robust focal length estimation. Bachelor's thesis, Czech Technical University in Prague, 2017. 4

[33] Torsten Sattler, Tobias Weyand, Bastian Leibe, and Leif Kobbelt. Image Retrieval for Image-Based Localization Revisited. In *British Machine Vision Conference (BMCV)*, 2012. 6

[34] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. Efficient & effective prioritized matching for large-scale image-based localization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 39(9): 1744–1756, 2017. 1

[35] Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Fredrik Kahl, and Tomas Pajdla. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 6

[36] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 6, 7

[37] Thomas Schops, Johannes L Schonberger, Silvano Galliani, Torsten Sattler, Konrad Schindler, Marc Pollefeys, and Andreas Geiger. A multi-view stereo benchmark with high-resolution images and multi-camera videos. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3260–3269, 2017. 7, 8

[38] Noah Snavely, Steven M Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *IJCV*, 80(2): 189–210, 2008. 1

[39] Jakub Sochor, Roman Juránek, Jakub Špaňhel, Lukáš Maršík, Adam Širokỳ, Adam Herout, and Pavel Zemčík. Comprehensive data set for automatic single camera visual speed measurement. *IEEE Transactions on Intelligent Transportation Systems*, 20(5):1633–1643, 2018. 1

[40] Henrik Stewénius, David Nistér, Fredrik Kahl, and Frederik Schaffalitzky. A minimal solution for relative pose with unknown focal length. *Image and Vision Computing*, 26(7): 871–877, 2008. 2, 7, 8

[41] Peter Sturm. On focal length calibration from two views. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, pages 145–150. IEEE, 2001. 2, 7, 8

[42] Peter Sturm, ZL Cheng, Peter CY Chen, and Aun Neow Poo. Focal length calibration from two views: method and analysis of singular cases. *Computer Vision and Image Understanding*, 99(1):58–95, 2005. 2

[43] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and Xiaowei Zhou. Loftr: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8922–8931, 2021. 7

[44] Magdalena Urbanek, Radu Horaud, and Peter Sturm. Combining off-and on-line calibration of a digital camera. In *Proceedings Third International Conference on 3-D Digital Imaging and Modeling*, pages 99–106. IEEE, 2001. 2

[45] Bin Xu and Zhenzhong Chen. Multi-level fusion based 3d object detection from monocular images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2345–2353, 2018. 1

[46] Zhengyou Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22(11):1330–1334, 2000. 1

[47] Zichao Zhang, Torsten Sattler, and Davide Scaramuzza. Reference pose generation for long-term visual localization via learned features and view synthesis. *International Journal of Computer Vision*, 129:821–844, 2021. 2, 7