# Supplementary Material - 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

## 1. Outline

Our supplementary material provides additional details on the method presented in Section 3 in the main paper, and promised experiments that support the material from the main paper. In Section 2, we summarize our algorithm for estimating focal lengths from a given fundamental matrix $\mathbf{F}$. In Section 3 we provide more details on the real focal length checking (RFC) described in Section 3.2 of the main paper along with additional experiments. To supplement the results presented in Section 5 of the main paper, we present experiments on real-world data incorporating different RANSAC variants and matches in Section 4. In Section 5 we provide an analysis of the relationship between the camera configuration and the convergence rates of our method. Additionally, in Section 6 we provide both synthetic and real-world experiments for the case when the focal lengths of the two cameras are equal.

## 2. Algorithm

The complete algorithm for our method for estimating focal lengths from a given fundamental matrix $\mathbf{F}$ is presented in Algorithm 1. The algorithm solves the constrained optimization problem described in Section 3 of the main paper by searching for stationary points of the Lagrange multiplier (Equation (3) in the main paper). In each iteration, this algorithm solves a system of two equations of degree four in two unknowns using the Gröbner basis method [15] (for more details see the main paper).

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 $\mathbf{K}_2$, the matrix $\mathbf{K}_1^\top \mathbf{F} \mathbf{K}_2$ is a valid essential matrix in each iteration.

For the case of equal focal lengths, *i.e.*, $f_1 = f_2$, the algorithm has the same structure as Algorithm 1 and it also solves the system of two equations of degree four in two unknowns. The only difference is that in each step $i = 1$.

**Algorithm 1**

**Input:** Fundamental matrix $\mathtt{F}$ ,
Priors $f_i^p, \mathbf{c}_i^p,\ i = 1, 2$,
$\epsilon_1$, *maxiter*
**Output:** $(f_1^\star, f_2^\star, \mathbf{c}_1^\star, \mathbf{c}_2^\star)$

*found* $\leftarrow 0$, $k \leftarrow 1$
$f_i^0 \leftarrow f_i^p, \mathbf{c}_i^0 \leftarrow \mathbf{c}_i^p, i = 1, 2$
$s^0 \leftarrow \langle f_1^0, f_2^0, \mathbf{c}_1^0, \mathbf{c}_2^0 \rangle$
**while** (**not** *found*) **and** ($k \leq$ *maxiter*) **do**
  $\Delta f_i \leftarrow \frac{1}{w_i^f}(\lambda_1 \frac{\partial \kappa_1}{\partial f_i}(s^{k-1}) + \lambda_2 \frac{\partial \kappa_2}{\partial f_i}(s^{k-1}))$
  $\Delta \mathbf{c}_i \leftarrow \frac{1}{w_i^{\mathbf{c}}}(\lambda_1 (\frac{\partial \kappa_1}{\partial \mathbf{c}_i}(s^{k-1}))^\top + \lambda_2 (\frac{\partial \kappa_2}{\partial \mathbf{c}_i}(s^{k-1}))^\top)$.
  $\lambda_1^k, \lambda_2^k \leftarrow$ The solution to two equations $\kappa_1^k = 0$ and $\kappa_2^k = 0$, which among all 16 solutions minimizes $|\lambda_1| + |\lambda_2|$
  $\Delta f_i^k \leftarrow \Delta f_i(\lambda_1^k, \lambda_2^k),\ \ i = 1, 2$
  $\Delta \mathbf{c}_i^k \leftarrow \Delta \mathbf{c}_i(\lambda_1^k, \lambda_2^k),\ \ i = 1, 2$
  $e^k \leftarrow \sum_{i=1,2} w_i^f (\Delta f_i^k)^2 + w_i^{\mathbf{c}} (\Delta \mathbf{c}_i^k)^\top \Delta \mathbf{c}_i^k$
  $f_i^k \leftarrow \Delta f_i^k + f_i^p, \quad \mathbf{c}_i^k = \Delta \mathbf{c}_i^k + \mathbf{c}_i^p, \quad i = 1, 2$
  **if** ($k > 1$ **and** $\frac{|e^k - e^{k-1}|}{e^k} < \epsilon_1$) **then**
    $\langle f_1, f_2, \mathbf{c}_1, \mathbf{c}_2 \rangle \leftarrow \langle f_1^k, f_2^k, \mathbf{c}_1^k, \mathbf{c}_2^k \rangle$
    *found* $\leftarrow 1$
  **else**
    $k \leftarrow k + 1$
  **end if**
**end while**
**if not** *found* **then**
  $\langle f_1, f_2, \mathbf{c}_1, \mathbf{c}_2 \rangle \leftarrow \langle f_1^{k-1}, f_2^{k-1}, \mathbf{c}_1^{k-1}, \mathbf{c}_2^{k-1} \rangle$
**end if**

## 3. Real Focal Length Checking

In this section, we provide further details on the real focal length checking (RFC) described in section 3.2 of the main paper.

$$f_1^2 = \frac{-\mathbf{F}_{33}\left(\mathbf{F}_{12}\mathbf{F}_{13}\mathbf{F}_{33}-\mathbf{F}_{13}^2\mathbf{F}_{32}+\mathbf{F}_{22}\mathbf{F}_{23}\mathbf{F}_{33}-\mathbf{F}_{23}^2\mathbf{F}_{32}\right)}{\mathbf{F}_{11}\mathbf{F}_{12}\mathbf{F}_{31}\mathbf{F}_{33}-\mathbf{F}_{11}\mathbf{F}_{13}\mathbf{F}_{31}\mathbf{F}_{32}+\mathbf{F}_{12}^2\mathbf{F}_{32}\mathbf{F}_{33}-\mathbf{F}_{12}\mathbf{F}_{13}\mathbf{F}_{32}^2+\mathbf{F}_{21}\mathbf{F}_{22}\mathbf{F}_{31}\mathbf{F}_{33}-\mathbf{F}_{21}\mathbf{F}_{23}\mathbf{F}_{31}\mathbf{F}_{32}+\mathbf{F}_{22}^2\mathbf{F}_{32}\mathbf{F}_{33}-\mathbf{F}_{22}\mathbf{F}_{23}\mathbf{F}_{32}^2} \tag{1}$$

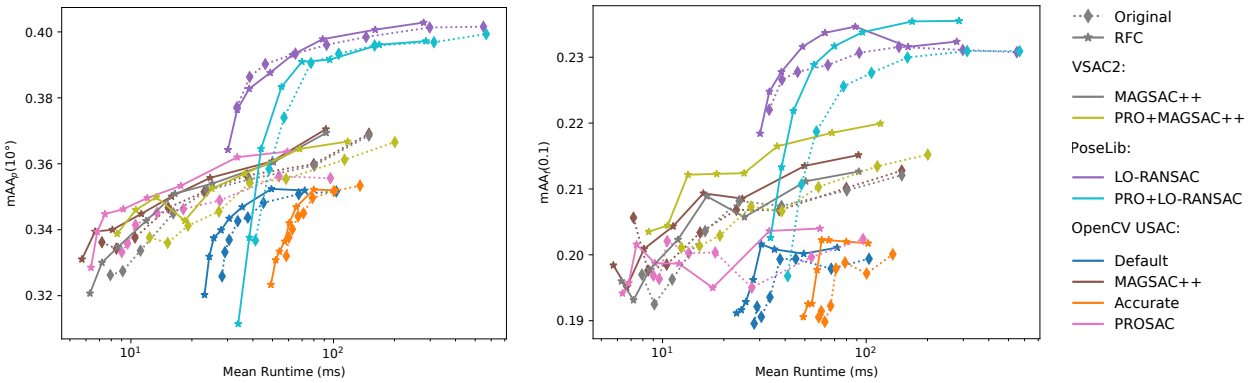



Figure 1. Figure shows the pose (left) and focal length (right) mean average accuracy on the Phototourism dataset [13] obtained after decomposing fundamental matrices with the Bougnoux formula [3]. To check the speed-accuracy trade-off we vary the total number of iterations of various RANSAC implementations with and without performing RFC.

We implemented RFC by utilizing formula (1) for computing the focal length from the elements of the given fundamental matrix $\mathbf{F}$. This formula was presented in [18]. To obtain $f_2^2$, the same formula can be used with $\mathbf{F}$ transposed. In the code we only check the signs of the numerator and denominator to conclude whether $f_i^2 > 0$. The check thus uses only computationally efficient operations. Note that the formula (1) assumes that both principal points lie at the origin of the camera coordinate system. When the positions of both principal points are known but different from the origin, the fundamental matrix can be easily transformed so that the principal points lie at the origin [23].

We implemented RFC into three state-of-the-art libraries for fundamental matrix estimation: OpenCV [4], VSAC [12] and PoseLib [14]. The implementations are modular, and thus allow us to test RFC with different types of RANSACs even within a single library.

In OpenCV we test four configurations:

- Default - LO-RANSAC [6],
- Accurate - LO-RANSAC [6] + GC-RANSAC [1],
- MAGSAC++ [2],
- PROSAC [5].

In VSAC we test the two configurations which both use MAGSAC++, one with PROSAC and one without it. All of the OpenCV and VSAC variants also utilize the oriented epipolar constraint [7] and DEGENSAC [8] degeneracy checks and SPRT verification [17]. In PoseLib we test the default configuration implementing LO-RANSAC with and without PROSAC. The modified versions of these libraries are available online.[1]

[1] https://github.com/kocurvik/robust_self_calibration

### 3.1. Speed-accuracy Trade-off

In Fig. 3 of the main paper we have shown that employing RFC within different implementations of RANSAC leads to improvement in both the speed of computation as well as the estimated poses when using the ground truth focal lengths to decompose the obtained fundamental matrices. Here, we show that this is also generally the case when decomposing the fundamental matrices using the Bougnoux formula [3] (Fig. 1) and our method (Fig. 2). We show that the accuracy and computation speed improvements can, in general, be observed both in terms of pose accuracy and estimated focal length accuracy.

We measured the run-time of each RANSAC variant on the Phototourism dataset [13] using LoFTR [24] correspondences. We set the epipolar threshold for all variants to 3 px. We ran the evaluation on the 200 sample pairs from each of the 12 test scenes of Phototourism. The pairs were randomly selected based on the co-visibility criteria as outlined in [13]. We let each of the RANSAC variants run for 50, 100, 500, 1000, 2000, 5000 and 10000 iterations. We prevented early termination by setting the confidence parameter to 1.0 for all methods. All tests were performed on an Intel i7-11800H CPU. For comparison, we averaged the wall-clock runtimes over all of the evaluated samples.

## 4. Additional Real-world Experiments

In this section, we present additional results on the Phototourism [13] and Aachen Day-Night v1.1 [25] datasets, as promised in Section 5 in the main paper. We test how the compared methods perform when we use different robust

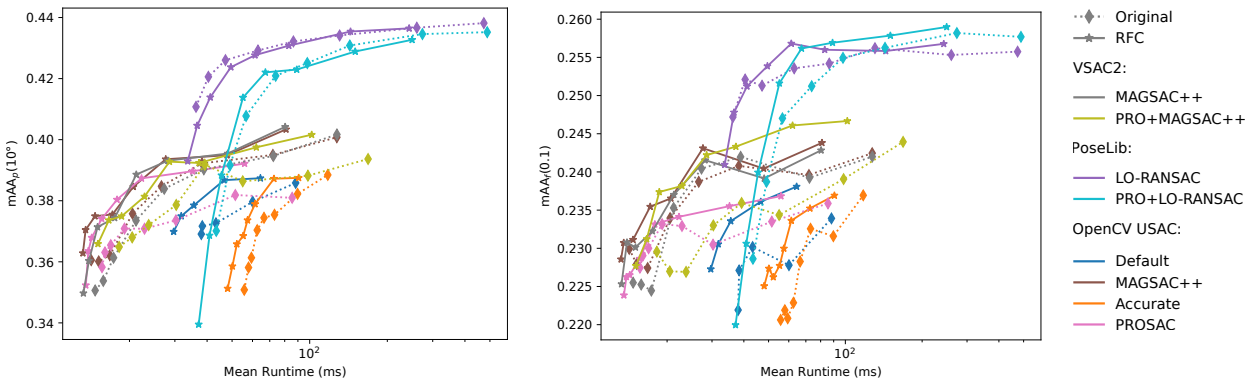


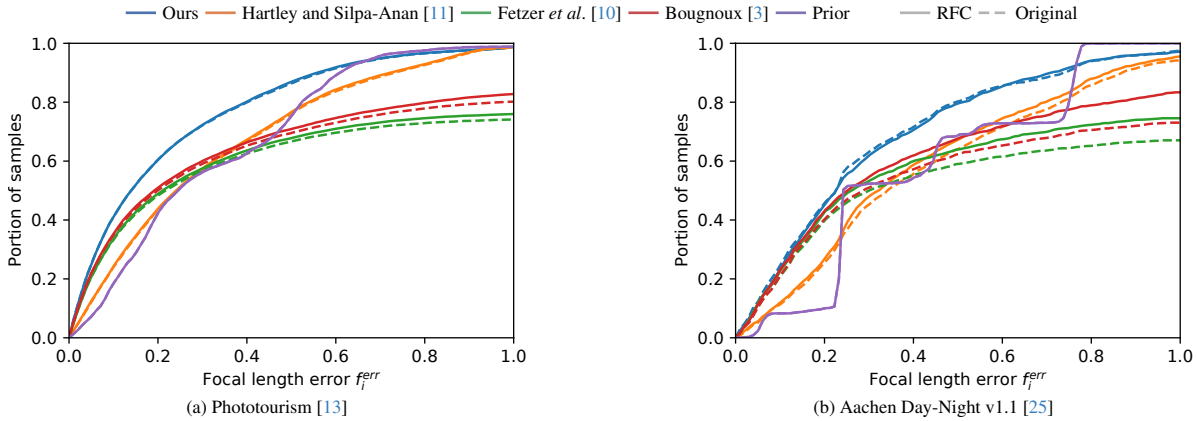

Figure 2. Figure shows the pose (left) and focal length (right) mean average accuracy on the Phototourism dataset [13] obtained after decomposing fundamental matrices with our method. To check the speed-accuracy trade-off we vary the total number of iterations of various RANSAC implementations with and without performing RFC.

| Method | RFC | Phototourism [13] | | | | | | Aachen Day-Night v1.1 [25] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Median | $\text{mAA}_p$ | | Median | $\text{mAA}_f$ | | Median | $\text{mAA}_p$ | | Median | $\text{mAA}_f$ | |
| | | $p_{err}$ | 10° | 20° | $f_{err}$ | 0.1 | 0.2 | $p_{err}$ | 10° | 20° | $f_{err}$ | 0.1 | 0.2 |
| Ours | | **5.63°** | 43.95 | **58.92** | **0.140** | **25.59** | **39.06** | 9.07° | 29.07 | 47.05 | 0.227 | **13.69** | **25.15** |
| | ✓ | **5.63°** | **44.04** | 59.17 | **0.140** | 25.54 | 39.03 | **8.85°** | **29.11** | **47.96** | **0.226** | 12.67 | 24.23 |
| Hartley [11] | | 8.74° | 31.45 | 48.38 | 0.244 | 13.23 | 24.14 | 11.87° | 21.49 | 39.64 | 0.341 | 6.17 | 12.44 |
| | ✓ | 8.71° | 31.52 | 48.62 | 0.241 | 13.46 | 24.44 | 11.69° | 21.80 | 40.77 | 0.321 | 6.46 | 13.06 |
| Fetzer [10] | | 8.10° | 36.06 | 49.88 | 0.217 | 20.87 | 31.64 | 12.87° | 24.46 | 38.96 | 0.300 | 11.25 | 21.42 |
| | ✓ | 7.94° | 36.23 | 50.24 | 0.211 | 20.95 | 31.95 | 11.21° | 25.57 | 41.70 | 0.261 | 12.33 | 22.99 |
| Bougnoux [3] | | 6.63° | 40.41 | 55.70 | 0.199 | 22.01 | 33.07 | 10.18° | 27.42 | 45.47 | 0.288 | 11.13 | 21.66 |
| | ✓ | 6.63° | 40.45 | 55.62 | 0.192 | 22.16 | 33.51 | 9.58° | 27.93 | 46.49 | 0.251 | 12.25 | 23.26 |
| Prior | | 11.03° | 22.78 | 42.42 | 0.246 | 8.83 | 19.81 | 13.08° | 17.15 | 37.28 | 0.242 | 4.11 | 6.55 |
| | ✓ | 11.01° | 22.81 | 42.53 | 0.246 | 8.83 | 19.81 | 12.52° | 17.63 | 38.71 | 0.242 | 4.11 | 6.55 |
| GT intrinsics | | 2.31° | 63.80 | 73.86 | — | — | — | 4.19° | 52.05 | 63.87 | — | — | — |
| | ✓ | 2.34° | 63.88 | 74.05 | | | | 4.02° | 53.21 | 65.89 | | | |

Table 1. Median errors for poses ($p_{err}$) and focal lengths ($f_i^{err}$) and mean average accuracy scores for poses ($\text{mAA}_p$) and estimated focal lengths ($\text{mAA}_f$) on 12 scenes from the Phototourism dataset [13] and the Aachen Day-Night v1.1 dataset [25]. Matches were produced using LoFTR [24] and the fundamental matrices were estimated using the PoseLib [14] implementation of LO-RANSAC [6]. RFC denotes real focal length checking.

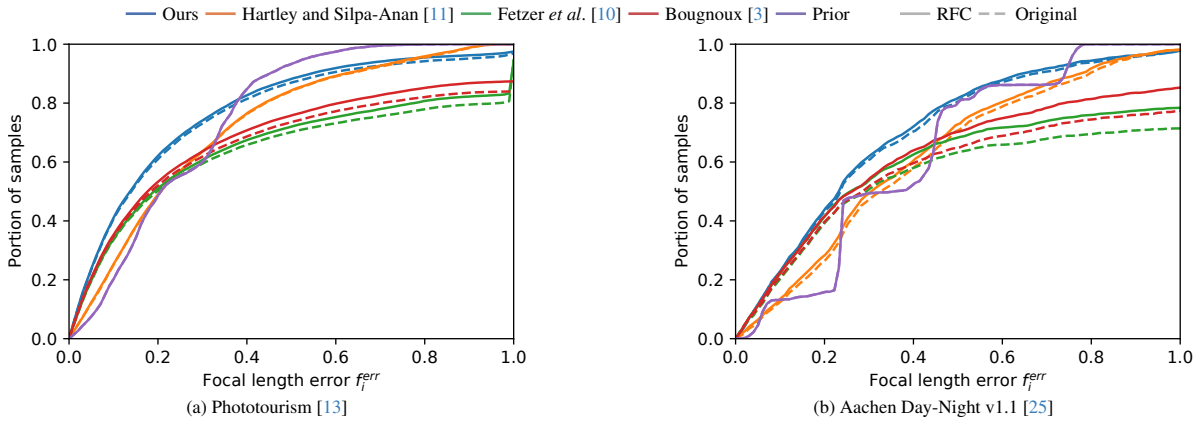
| Method | RFC | Phototourism [13] | | | | | | Aachen Day-Night v1.1 [25] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Median | $\text{mAA}_p$ | | Median | $\text{mAA}_f$ | | Median | $\text{mAA}_p$ | | Median | $\text{mAA}_f$ | |
| | | $p_{err}$ | 10° | 20° | $f_{err}$ | 0.1 | 0.2 | $p_{err}$ | 10° | 20° | $f_{err}$ | 0.1 | 0.2 |
| Ours | | 6.47° | 40.19 | 56.99 | 0.144 | 24.00 | 37.91 | 10.01° | 25.93 | 45.59 | 0.235 | 12.66 | 23.30 |
| | ✓ | **6.31°** | **40.63** | **57.40** | **0.143** | **24.27** | **38.20** | **8.91°** | **27.73** | **47.84** | **0.232** | **12.67** | **23.58** |
| Hartley [11] | | 8.90° | 30.83 | 48.49 | 0.240 | 12.81 | 23.94 | 12.01° | 20.13 | 39.20 | 0.326 | 6.59 | 13.376 |
| | ✓ | 8.78° | 31.00 | 48.60 | 0.239 | 12.97 | 24.09 | 11.40° | 21.13 | 40.68 | 0.304 | 7.28 | 14.53 |
| Fetzer [10] | | 9.15° | 32.51 | 47.94 | 0.217 | 19.87 | 30.90 | 12.42° | 22.23 | 38.76 | 0.294 | 11.27 | 20.98 |
| | ✓ | 8.80° | 33.08 | 48.72 | 0.210 | 20.03 | 31.24 | 11.17° | 23.20 | 41.28 | 0.259 | 12.31 | 22.55 |
| Bougnoux [3] | | 7.53° | 36.48 | 53.43 | 0.201 | 20.58 | 31.94 | 10.20° | 25.75 | 44.86 | 0.285 | 11.63 | 21.39 |
| | ✓ | 7.40° | 36.77 | 53.59 | 0.192 | 20.91 | 32.54 | 9.88° | 25.92 | 45.13 | 0.262 | 12.40 | 22.66 |
| Prior | | 10.74° | 23.37 | 43.46 | 0.246 | 8.83 | 19.82 | 11.93° | 17.82 | 39.47 | 0.355 | 6.78 | 10.60 |
| | ✓ | 10.68° | 23.60 | 43.62 | 0.246 | 8.83 | 19.82 | 11.48° | 18.62 | 40.83 | 0.355 | 6.78 | 10.60 |
| GT intrinsics | | 3.02° | 59.58 | 72.07 | — | — | — | 4.77° | 48.98 | 64.11 | — | — | — |
| | ✓ | 3.00° | 59.78 | 72.27 | | | | 4.50° | 50.55 | 65.50 | | | |

Table 2. Median errors for poses ($p_{err}$) and focal lengths ($f_i^{err}$) and mean average accuracy scores for poses ($\text{mAA}_p$) and estimated focal lengths ($\text{mAA}_f$) on 12 scenes from the Phototourism dataset [13] and the Aachen Day-Night v1.1 dataset [25]. Matches were produced using SuperPoint [9] in combination with SuperGlue [19]. The OpenCV [4] implementation of MAGSAC++ [2] was used to estimate the fundamental matrices. RFC denotes the real focal length checking.




(a) Phototourism [13] (b) Aachen Day-Night v1.1 [25]

Figure 3. Plots showing the portion of samples for which the estimated focal lengths were below a given $f_i^{err}$ threshold. Both cameras are assumed to have different unknown focal lengths. To obtain correspondences we used LoFTR [24]. We produced the fundamental matrices using the PoseLib [14] implementation of LO-RANSAC [6].




(a) Phototourism [13] (b) Aachen Day-Night v1.1 [25]

Figure 4. Plots showing the portion of samples for which the estimated focal lengths were below a given $f_i^{err}$ threshold. Both cameras are assumed to have different unknown focal lengths. To obtain correspondences we used SuperPoint [9] and SuperGlue [19]. We produced the fundamental matrices using the OpenCV [4] implementation of MAGSAC++ [2]

.

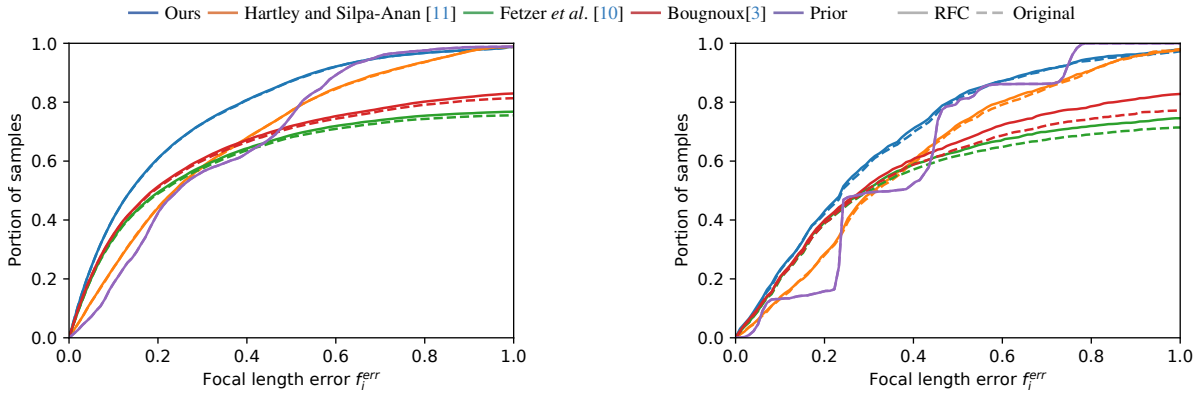
estimators of the fundamental matrix as well as different matches.

In the main paper in Table 1 and Figs. 4a and 4b we present experiments with LoFTR [24] matches. We also performed experiments using the combination of the SuperPoint [9] keypoint detector along with the SuperGlue [19] matching network to obtain matches in images. Before inference, we resize the images such that the larger dimension of the image is 2048 pixels. Note that the memory requirements of the SuperGlue network depend on the number of detected keypoints. This led to insufficient GPU memory for some of the pairs in both datasets. We therefore skip such pairs, but add additional pairs to keep 1000 sample pairs per each scene.

To estimate the fundamental matrices in the main paper, we use MAGSAC++ [2] as implemented in OpenCV [4]. We have set the epipolar threshold to 3 px using the validation set. Here we also include experiments using LO-RANSAC [6] implemented in the PoseLib library [14]. We also use it with the epipolar threshold set to 3 px based on performance on the validation set. We set the number of iterations to 10000.

Table 1 and Fig. 3 show the results on both datasets for the combination of LoFTR matches with PoseLib LO-RANSAC. We present the results for the combination of SuperPoint, SuperGlue and MAGSAC++ in Table 2 and Fig. 4. The results for the same matches in combination with PoseLib LO-RANSAC are shown in Table 3 and Fig. 5.

These results show that our method outperforms competing approaches when different RANSAC and matching methods are considered in the fundamental matrix estimation pipeline. The results also show that performing the real focal length check (RFC) generally leads to better accuracy in both the estimated poses and focal lengths across




(a) Phototourism [13]

(b) Aachen Day-Night v1.1 [25]

Figure 5. Plots showing the portion of samples for which the estimated focal lengths were below a given $f_i^{err}$ threshold. Both cameras are assumed to have different unknown focal lengths. To obtain correspondences we used SuperPoint [9] in combination with SuperGlue [19]. We produced the fundamental matrices using the PoseLib [14] implementation of LO-RANSAC [6]

.

| Method | RFC | Phototourism [13] | | | | | | Aachen Day-Night v1.1 [25] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Median | mAA$_p$ | | Median | mAA$_f$ | | Median | mAA$_p$ | | Median | mAA$_f$ | |
| | | $p_{err}$ | 10° | 20° | $f_{err}$ | 0.1 | 0.2 | $p_{err}$ | 10° | 20° | $f_{err}$ | 0.1 | 0.2 |
| Ours | | 5.92° | 42.75 | 59.25 | **0.139** | **24.85** | 38.77 | 10.14° | 26.40 | 45.95 | 0.240 | 12.07 | 22.90 |
| | ✓ | **5.89°** | **42.80** | **59.33** | **0.139** | 24.83 | **38.82** | **9.69°** | **27.03** | **46.91** | **0.237** | **12.61** | **23.25** |
| Hartley [11] | | 8.45° | 31.92 | 49.55 | 0.238 | 13.12 | 24.32 | 11.82° | 21.09 | 40.07 | 0.317 | 6.93 | 14.00 |
| | ✓ | 8.51° | 31.87 | 49.57 | 0.237 | 13.24 | 24.30 | 11.46° | 21.42 | 40.84 | 0.312 | 7.11 | 14.14 |
| Fetzer [10] | | 8.35° | 34.65 | 49.71 | 0.211 | 20.64 | 31.82 | 12.57° | 22.09 | 38.40 | 0.302 | 10.41 | 20.34 |
| | ✓ | 8.35° | 34.66 | 49.91 | 0.206 | 20.79 | 32.06 | 12.45° | 22.39 | 39.04 | 0.291 | 10.65 | 20.78 |
| Bougnoux [3] | | 6.90° | 38.88 | 55.51 | 0.191 | 21.54 | 33.11 | 10.20° | 25.49 | 44.67 | 0.295 | 10.82 | 20.63 |
| | ✓ | 6.93° | 38.76 | 55.35 | 0.188 | 21.61 | 33.27 | 10.13° | 25.55 | 44.48 | 0.285 | 11.15 | 21.25 |
| Prior | | 10.64° | 23.60 | 43.80 | 0.246 | 8.83 | 19.82 | 11.70° | 17.88 | 40.24 | 0.355 | 6.78 | 10.60 |
| | ✓ | 10.68° | 23.57 | 43.78 | 0.246 | 8.83 | 19.82 | 11.57° | 17.97 | 40.45 | 0.355 | 6.78 | 10.60 |
| GT intrinsics | | 2.66° | 62.56 | 74.50 | — | — | — | 4.32° | 50.83 | 65.57 | — | — | — |
| | ✓ | 2.67° | 62.62 | 74.54 | | | | 4.26° | 51.35 | 66.03 | | | |

Table 3. 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 [13] and the Aachen Day-Night v1.1 dataset [25]. Matches were produced using SuperPoint [9] in combination with SuperGlue [19]. The PoseLib [14] implementation of LO-RANSAC [6] was used to estimate the fundamental matrices. RFC denotes the real focal length checking.

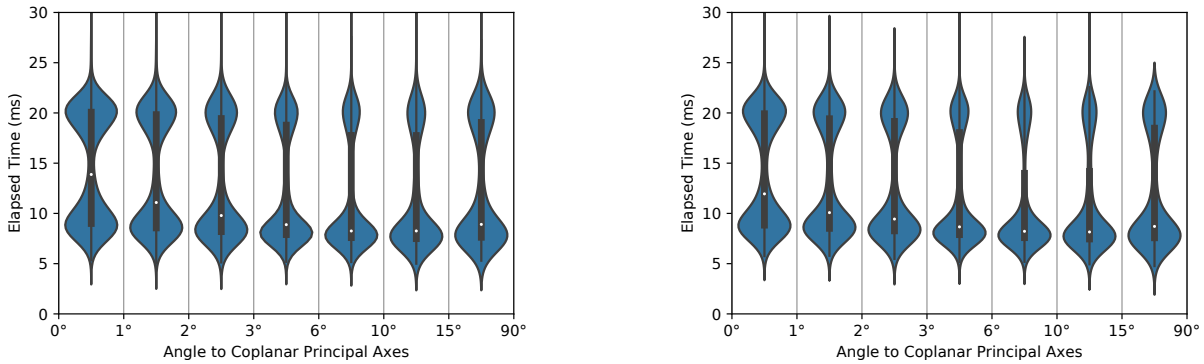
the compared methods.

## 5. Effect of Camera Configuration on Convergence

In this section, we provide further insight into the convergence of our method on real data. Fig. 6 shows the distributions of convergence times expressed in wall-clock runtime for different camera configurations on the Phototourism dataset [13]. We express the camera configuration as the angle required to turn one of the cameras so that their principal axes would be coplanar thus resulting in the degenerate configuration. To determine the angle, we used the poses from the COLMAP [20] 3D reconstructions that are provided by the dataset authors as the ground truth. The distributions for all camera configurations have two peaks. The bottom peak corresponds to termination within afew iterations, and the top peak corresponds to termination after the maximum number of iterations is reached. For the configurations that have principal axes close to coplanar (degenerate), it can be seen that reaching the maximum number of iterations is more likely compared to other (non-degenerate) configurations. The runtimes are thus longer on average near those configurations.

## 6. Cameras with Equal Focal Lengths

In this section, we present additional details and experiments for the case when the focal lengths of the two cameras are assumed to be equal. We use the algorithm modified for this scenario as described in Section 2.




(a) Without RFC (b) With RFC

Figure 6. The violin plots show the wallclock runtimes of our method on the Phototourism dataset [13] when different camera configurations are considered. The plot shows that when the camera principal axes are close to coplanar (intersecting) our method is more likely to terminate when the maximum number of iterations is reached.

### 6.1. Synthetic Experiments

We perform the same synthetic experiments as for the case of two cameras with unknown but different focal lengths (*cf*. Section 4 in the main paper). We use the same notation for camera configurations and added noise. We note that the configuration of $\mathcal{C}(0°, 0)$ has the two principal axes meeting at a point equidistant from both camera centers. Thus it is a degenerate configuration leading to a generic singularity in the Kruppa equations [16] when the focal lengths are considered equal.

**Convergence:** Figures 7a and 7b show convergence rates in non-degenerate and degenerate configurations respectively. The method converges in the majority of cases in a few iterations, even when strict thresholds are considered. The algorithm also converges within 50 iterations for the majority of the cases on real data from the ETH3D Multi-view datset [21] as shown in Fig. 7c. Therefore, we set the maximum number of iterations to 50 in further experiments.

**Accuracy of the Estimated Focal Length:** For comparison, we use the method by Fetzer *et al*. [10] to optimize for a single focal length for both cameras. Instead of the Bougnoux formula [3], we use the formula proposed by Sturm [23] that is specific to the case of equal focal lengths. We use Hartley and Silpa-Anan's method [11] as usual, but on output we average the two focal lengths produced. We have also implemented Hartley and Silpa-Anan's method to estimate a single focal length while utilizing Sturm's formula in the iterative optimization loop, but we found it to produce less stable results than those for the selected approach.

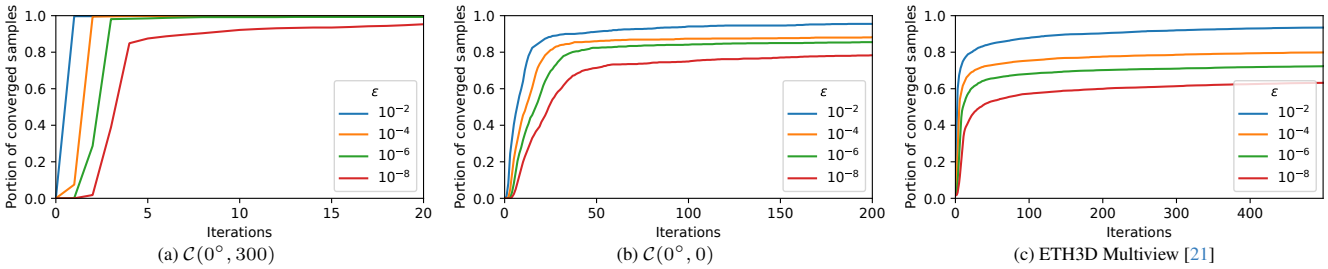
We also evaluate the minimal 6-point algorithm [22] implemented within LO-RANSAC [6, 14]. We set the number of iterations to 10000 and the epipolar threshold to 1 px.

Fig. 8 shows the results of the synthetic experiments. We observe a similar behavior as in synthetic experiments for the case of different focal lengths presented in the main paper (*cf*. Section 4). However, there are some differences that are specific for the equal focal length case and different methods tested in this case. Hartley and Silpa-Anan's iterative method performs poorly in this scenario. Sturm's method for equal focal lengths is more stable than the Bougnoux formula for the general case. The method by Fetzer *et al*. provides results similar to those of Sturm's formula, but it is less stable when bad initialization is used or when the cameras are near the degenerate configuration. The 6-point minimal algorithm provides stable results even in proximity of the degenerate configuration. Our method performs on par with the minimal 6-point algorithm, thus providing a viable alternative.

### 6.2. Real-world Experiments

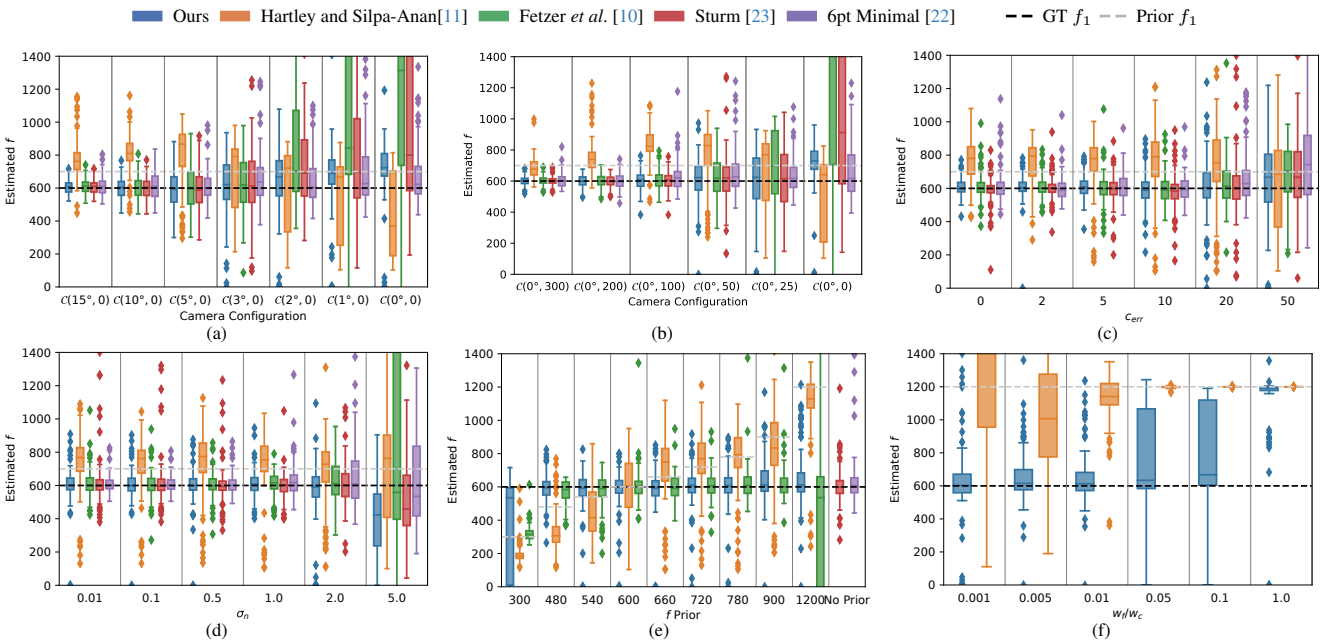
In the main paper (*cf*. Sec 5, Table 3), we have presented a comparison of various methods on the ETH3D Multi-view dataset [21]. Here, we present additional experiments which include evaluation of the compared methods when RFC is used. We also include experiments using a different RANSAC variant. We use LoFTR [24] to obtain matches. We matched all pairs of images that had co-visibility factor [13] greater than 0.1. Prior to inference, we resized the images so that the larger size is 1024 px. We estimated the fundamental matrices using LO-RANSAC [6] implemented in PoseLib [14] and MAGSAC++ [2] implemented in OpenCV [4].

The results show that using our method in conjunction with RFC leads to good accuracy in terms of both the estimated poses and focal lengths. In comparison, Sturm's method achieves the highest $\text{mAA}_p$ scores, but performs poorly with respect to focal length accuracy. On the other hand, the minimal solver shows good results in terms of




(a) $\mathcal{C}(0^\circ, 300)$ (b) $\mathcal{C}(0^\circ, 0)$ (c) ETH3D Multiview [21]

Figure 7. 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 prior as $f^p = 660$. For (c) we used the ETH3D Multiview dataset [21].




(a) (b) (c) (d) (e) (f)

Figure 8. Synthetic experiments: Box plots for the estimated focal length equal for the two cameras. Comparison of the methods as (a, b) the camera configuration approaches the degenerate configuration, (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 $f$, (f) we vary the relative weights of the focal length and principal point priors. We use priors $f^p = 700$ for (a, b, c, d), $f^p = 1200$ for (f), $\sigma_n = 1$ for (a, b, c, e, f), $\sigma_p = 10$ for (a, b, d, e, f). For (c, d, e, f) we randomly sample the configuration $\mathcal{C}(\theta, y)$ with $\theta \in [-15^\circ, 15^\circ]$ and $y \in [-200, 200]$.

the estimated focal lengths but shows weak results when it comes to pose accuracy. Our method is thus a viable alternative to other existing approaches, when both cameras are assumed to have equal focal lengths.

## References


[1] Daniel Barath and Jiří Matas. Graph-cut ransac. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6733–6741, 2018. 2

[2] 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, 3, 4, 6, 8

[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, 5, 6

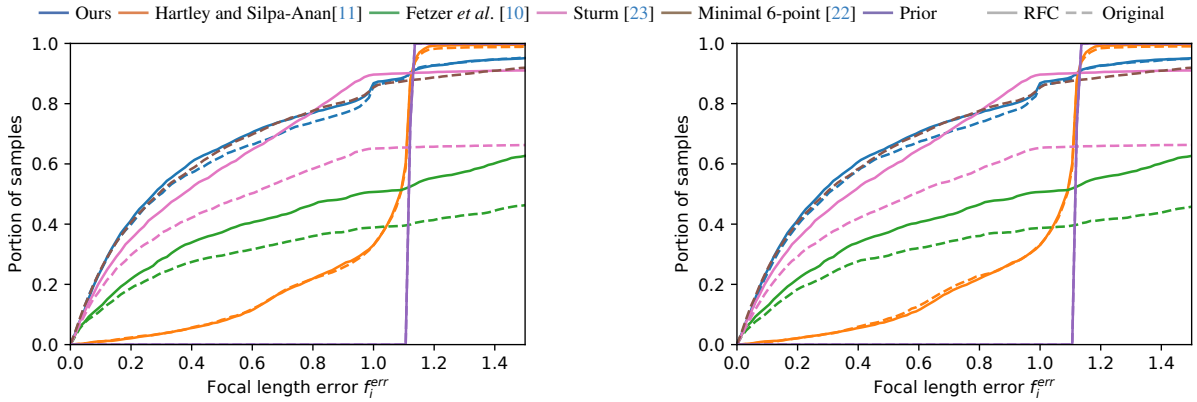
[4] Gary Bradski. The opencv library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11):120–123, 2000. 2, 3, 4, 6, 8

[5] 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

[6] 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. 2, 3, 4, 5, 6,

| Method | RFC | PoseLib LO-RANSAC [6, 14] | | | | | | OpenCV MAGSAC++ [2, 4] | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 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 | | **15.20°** | 17.69 | 33.66 | 0.298 | 14.23 | 24.12 | 16.51° | 16.70 | 32.24 | 0.299 | 12.96 | 23.10 |
| | ✓ | 15.70° | 16.97 | 33.01 | **0.274** | 13.82 | 24.26 | **15.70°** | 16.97 | 33.01 | **0.274** | 13.82 | 24.26 |
| Hartley [11] | | 21.09° | 12.41 | 24.40 | 0.523 | 1.089 | 0.50 | 1.11° | 12.38 | 24.26 | 1.090 | 0.45 | 0.99 |
| | ✓ | 21.11° | 12.24 | 24.19 | 1.088 | 0.57 | 1.06 | 21.70° | 12.34 | 24.14 | 1.088 | 0.57 | 1.06 |
| Fetzer [10] | | 23.18° | 18.56 | 29.04 | 1.663 | 7.43 | 11.53 | 23.93° | 17.64 | 27.93 | 1.728 | 6.76 | 10.95 |
| | ✓ | 20.27° | 17.86 | 29.65 | 0.950 | 8.19 | 13.10 | 19.72° | 17.89 | 30.08 | 0.950 | 8.19 | 13.10 |
| Sturm [23] | | 16.40° | **21.82** | **34.93** | 0.592 | 10.98 | 18.05 | 17.98° | **20.64** | **33.30** | 0.582 | 10.39 | 17.55 |
| | ✓ | 16.42° | 20.61 | 34.26 | 0.373 | 12.37 | 20.92 | 17.18° | 19.36 | 33.01 | 0.373 | 12.37 | 20.92 |
| Minimal 6pt [22] | — | 36.18° | 16.42 | 25.74 | 0.291 | **14.43** | **24.29** | 36.18° | 16.42 | 25.74 | 0.291 | **14.43** | **24.29** |
| Prior | | 24.38° | 11.25 | 20.80 | 1.119 | 0.00 | 0.00 | 24.60° | 11.24 | 20.92 | 1.119 | 0.00 | 0.00 |
| | ✓ | 24.53° | 11.11 | 20.74 | 1.119 | 0.00 | 0.00 | 24.45° | 11.28 | 21.07 | 1.119 | 0.00 | 0.00 |
| GT intrinsics | | 5.95° | 43.20 | 55.72 | — | — | — | 6.40° | 41.73 | 54.16 | — | — | — |
| | ✓ | 5.87° | 43.30 | 56.00 | | | | 6.41° | 41.54 | 54.57 | | | |

Table 4. Median errors for poses ($p_{err}$) and focal lengths ($f^{err}$) and mean average accuracy scores for poses (mAA$_p$) and estimated focal lengths (mAA$_f$) on the ETH3D Multiview dataset [21]. LoFTR [24] was used to produce matches. To estimate the fundamental matrices we used two different estimators: PoseLib LO-RANSAC [6, 14] and OpenCV MAGSAC++ [2, 4]. The minimal solver was implemented within the PoseLib LO-RANSAC framework. RFC denotes real focal length checking.




(a) PoseLib LO-RANSAC [6, 14]

(b) OpenCV Magsac++ [2, 4]

Figure 9. Plots showing the portion of samples for which the estimated focal lengths were below a given $f^{err}$ threshold on the ETH 3D Multview dataset [21]. The focal lengths for the two cameras are assumed to be equal. To obtain correspondences we used LoFTR [24]. We produced the fundamental matrices using (a) the PoseLib [14] implementation of LO-RANSAC [6] (b) the OpenCV [4] implementation of MAGSAC++ [2].


8

[7] Ondrej Chum, Tomás Werner, and Jiri Matas. Epipolar geometry estimation via ransac benefits from the oriented epipolar constraint. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, pages 112–115. IEEE, 2004. 2

[8] 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

[9] 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. 3, 4, 5

[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. 3, 4, 5, 6, 7, 8

[11] 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. 3, 4, 5, 6, 7, 8

[12] 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

[13] 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, 3, 4, 5, 6

[14] Viktor Larsson and contributors. PoseLib - Minimal Solvers for Camera Pose Estimation, 2020. 2, 3, 4, 5, 6, 8

[15] 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. 1

[16] 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. 6

[17] Jiri Matas and Ondrej Chum. Randomized ransac with sequential probability ratio test. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, pages 1727–1732. IEEE, 2005. 2

[18] Oleh Rybkin. Robust focal length estimation. Bachelor's thesis, Czech Technical University in Prague, 2017. 2

[19] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superglue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4938–4947, 2020. 3, 4, 5

[20] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 5

[21] 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. 6, 7, 8

[22] 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. 6, 7, 8

[23] 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, 6, 7, 8

[24] 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. 2, 3, 4, 6, 8

[25] 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, 3, 4, 5