# Noisy One-point Homographies are Surprisingly Good - Supplementary Material

Yaqing Ding[1,2], Jonathan Astermark[1], Magnus Oskarsson[1], and Viktor Larsson[1]
[1] Centre for Mathematical Sciences, Lund University
[2] Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
yaqing.ding@cvut.cz, {jonathan.astermark, magnus.oskarsson, viktor.larsson}@math.lth.se

## 1. Overview

In the supplementary we present the following
- Special case: Pure rotation (Section 2)
- Additional qualitative results (Section 3)
- More detailed results on HEB (Section 4)

## 2. Pure Rotation

In this section, we present two additional 1-point minimal solvers for the special case of pure rotation, both in the calibrated ($\mathbf{H} = \mathbf{R}$) and uncalibrated ($\mathbf{H} = \mathbf{K}\mathbf{R}\mathbf{K}^{-1}$) setting.

### 2.1. The Calibrated 1-SIFT Solver

First we consider the case of rotational motion with known intrinsic calibration. Given one point correspondence with known orientation under pure rotation, we have

$$\mathbf{x}_{12} = \mathbf{R}\mathbf{x}_{11}, \; \mathbf{l}_{12} = \mathbf{R}\mathbf{l}_{11}. \tag{1}$$

where we have leveraged the line constraint derived in the main paper. Note that $\mathbf{R}^{-T} = \mathbf{R}$ in this case. Then, we also have

$$\mathbf{x}_{12} \times \mathbf{l}_{12} = \mathbf{R}(\mathbf{x}_{11} \times \mathbf{l}_{11}). \tag{2}$$

By stacking the columns of (1) and (2), we obtain

$$\begin{aligned} &\mathbf{M}_2 = \mathbf{R}\mathbf{M}_1, \text{ with} \\ &\mathbf{M}_2 = [\mathbf{x}_{12}, \quad \mathbf{l}_{12}, \quad \mathbf{x}_{12} \times \mathbf{l}_{12}], \\ &\mathbf{M}_1 = [\mathbf{x}_{11}, \quad \mathbf{l}_{11}, \quad \mathbf{x}_{11} \times \mathbf{l}_{11}]. \end{aligned} \tag{3}$$

The unique solution for the rotation matrix is then given by $\mathbf{R} = \mathbf{M}_2\mathbf{M}_1^{-1}$.

### 2.2. The Unknown Focal Length Case

As shown in [6], for most modern CCD or CMOS cameras, it is often reasonable to assume that the cameras have square-shaped pixels, and the principal point coincides with the image center. In this case, if a pair of images are taken with a single moving camera with fixed intrinsic, the only unknown camera parameter is the constant but unknown focal length. Assuming the only unknown calibration parameter of the camera is the focal length, let $\mathbf{K}_2 = \mathbf{K}_1 = \mathbf{K}$.

Figure 1. **First Row:** Sample images from the Phone dataset used in Section 2.3. **Middle Row:** Illumination pair from the HPatches dataset. **Bottom Row:** Viewpoint pair from the HPatches dataset.

Assuming central principal point and zero skew, the intrinsic matrix can be formulated as $\mathbf{K} = \mathrm{diag}(f, f, 1)$, and we

| Methods | Timings(ms) | Inliers (%) | $\epsilon_f$(%) |
|---|---|---|---|
| 4-point [7] | 14.7 | 94.1 | 5.99 |
| 2-point [8] | 15.7 | 88.9 | - |
| 1-SIFT(rot) | 15.2 | 88.8 | - |
| 2-point($f$) [5] | 15.4 | 94.2 | 5.89 |
| 1-SIFT($f$) | 14.9 | 94.2 | **5.86** |

Table 1. Comparison of different solvers under pure rotation.

| Solver | mAA@5° | mAA@10° | mAA@2m | mAA@5m | Inliers | Runtime |
|---|---|---|---|---|---|---|
| 4-point [7] | 4.73 | 12.0 | 9.60 | 26.0 | 53.8 | 53.0 |
| 3-SIFT [1] | 7.44 | 18.6 | 10.4 | 26.0 | 64.4 | 80.3 |
| 2-SIFT [3] | 8.88 | 21.9 | 11.2 | 30.0 | 70.6 | 157.8 |
| 2-AC [2] | 10.9 | 26.6 | 13.0 | 33.1 | 71.8 | 55.8 |
| 2-SIFT (ours) | 8.64 | 20.3 | 10.9 | 27.9 | 67.7 | 81.1 |
| 1-SIFT (ours) | **16.4** | **37.0** | **14.0** | **36.1** | **81.2** | **45.8** |

Table 2. **Evaluation on the *Piazza del Popolo* scene without SNN ratio.**

| | | Viewpoint | | | Illumination | | | Full | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Solver | Constraint | 1px | 3px | 5px | 1px | 3px | 5px | 1px | 3px | 5px |
| 1-SIFT | Paper | 25.3 | 53.1 | 63.5 | 32.9 | 62.3 | 72.0 | 28.8 | 56.8 | 67.3 |
| 1-SIFT | SSO [9] | 25.6 | 53.0 | 63.7 | 32.1 | 62.2 | 72.3 | 28.6 | 56.7 | 67.5 |
| 1-SIFT | Random | 25.6 | 53.1 | 63.2 | 31.5 | 61.8 | 71.3 | 27.8 | 56.3 | 66.2 |

Table 3. **Ablation of the influence with different lines on `HPatches`.**

have

$$\sigma_1 \mathbf{K}^{-1}\mathbf{m}_{12} = \mathbf{R}\mathbf{K}^{-1}\mathbf{m}_{11}, \tag{4}$$

which is equivalent to

$$|\sigma_1 \mathbf{K}^{-1}\mathbf{m}_{12}| = |\mathbf{K}^{-1}\mathbf{m}_{11}|. \tag{5}$$

This directly gives a quadratic equation in $f$. In fact it will be linear in $f^2$ and since the focal length should be real and positive, there cannot be more than one solution. Once the focal length is known, the unique rotation matrix can be found with the method from Sec. 2.1.

### 2.3. Evaluation of Rotation Estimation

To show the benefits of the proposed solvers, we tested them on the data recorded under pure rotation from smartphone devices: the wide-angle cameras of the iPhone 11 and iPhone 13 with focal lengths of 29mm and 26mm, respectively. The sequences were captured at the resolution of 1280x720@30Hz with 103 332 image pairs in total. The ground truth focal lengths were obtained by offline calibration of the cameras using COLMAP [10]. For the calibrated case, we compare the proposed 1-SIFT solver with the 2-point minimal solver [8]. For the uncalibrated case, we compare the proposed 1-SIFT($f$) solver with the 2-point focal length solver [5]. All solvers use the 4-point solver for non-minimal fitting. The results for both calibrated case and unknown focal length cases are shown in Table 1.

## 3. Additional Qualitative Results

For a qualitative study, we compared our 1-SIFT solver to the traditional 4-point solver using the ground-truth image pairs and matches in `HEB`. For the 1-SIFT solver, we exhaustively used all inlier matches as minimal samples. For the 4-point solver, we randomly sampled as many minimal samples as there were matches, resulting in an equal number of minimal samples for both solvers. For each minimal sample, we estimated the homography and calculated the number of inliers based on the ground-truth matches, using a reprojection error of 20 pixels. For each solver, we plotted the 5 best homographies in terms of the number of inliers.

We also ran each solver in GC-RANSAC using *all* matches (not just ground-truth inliers) and plotted the output. The result on 3 different scenes in `HEB`was presented in the main text. In figure 2, we show some results from the remaining 7 scenes. We see that even if the 1-point solver gives noisy initial estimates, in many cases they are sufficiently good as initialization for the local optimization in GC-RANSAC. Meanwhile, in many cases the 4-point solver struggles to find any good suggestions at all, which can be explained by the low probability of sampling four good inlier points.

## 4. More Detailed Results on `HEB`

In Figures 3, 4, and 5, we show the the per-scene results of the `HEB` dataset. We can see that the proposed 1-SIFT solver achieves comparable accuracy to the state-of-the-art 2-AC solver while having lower runtime on each sequence. In Table 2, we show the results on the *Piazza del Popolo* scene without SNN ratio. In this case, the advantage of 1-point solver is more significant. In Table 3, we show the results by fixing the affine parameters from SIFT, and using different constraints.

## References


[1] Daniel Barath. Five-point fundamental matrix estimation for uncalibrated cameras. In *CVPR*, pages 235–243, 2018. 2

[2] Daniel Barath and Levente Hajder. Novel ways to estimate homography from local affine transformations. In *11th In-*

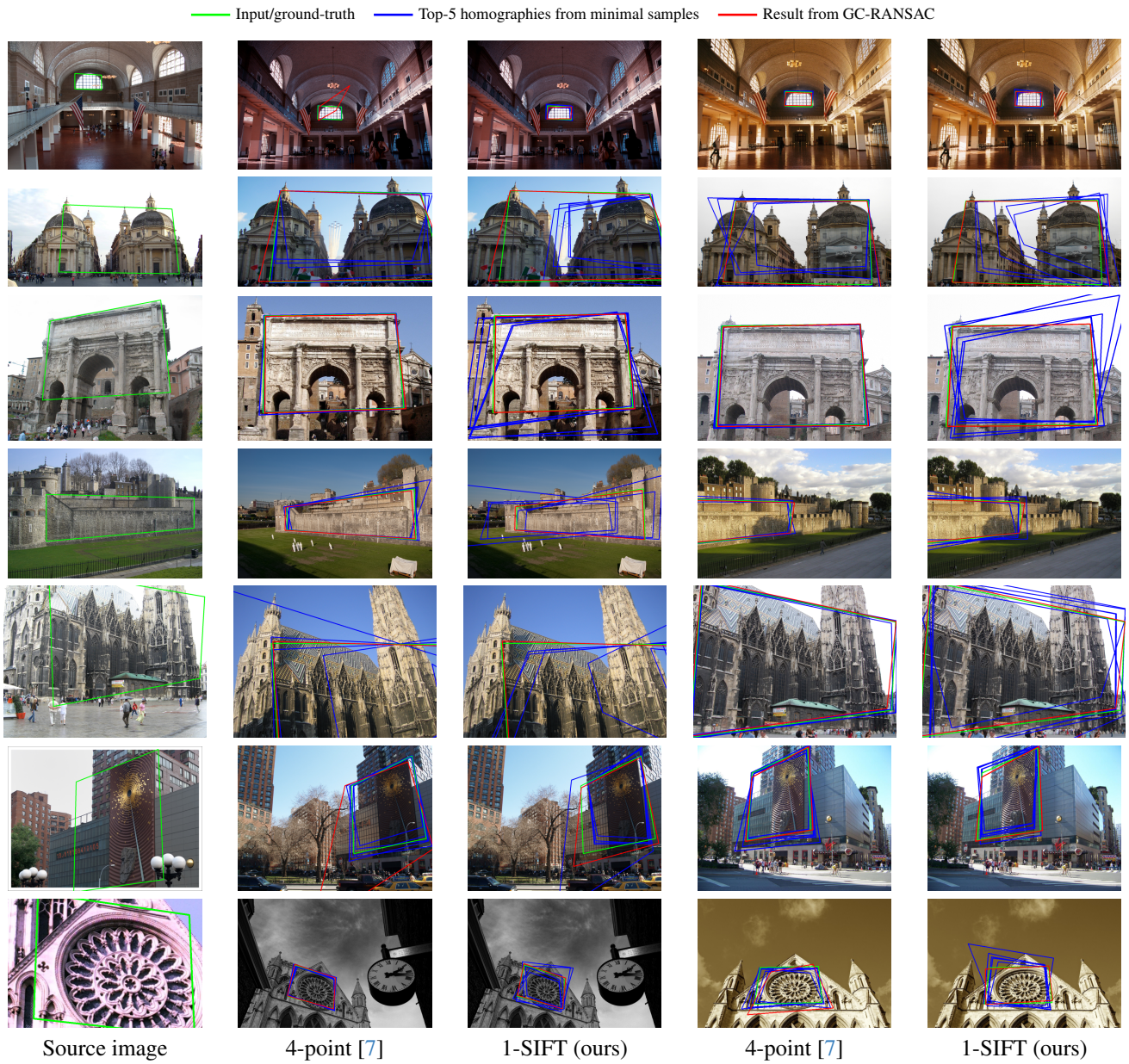



Figure 2. **Additional qualitative evaluation on HEB [4].** As a supplement to the main paper, we present qualitative results on the remaining 7 scenes from the HEB dataset. A plane from the source image is transformed to two other viewpoints by different homographies. The homographies are estimated using 4-point solver and 1-SIFT (ours). In blue are the best estimates on minimal samples taken from ground-truth inliers. In red are estimates using GC-RANSAC applied to all correspondences. Ground truths are shown in green. We see that even if the minimal estimates for our solver are noisy, they are often sufficient for good results in GC-RANSAC. Meanwhile, the 4-point solver struggles for planes with low inlier ratios. Some images have been cropped for clearer visualization, but the algorithms were run on the full images.

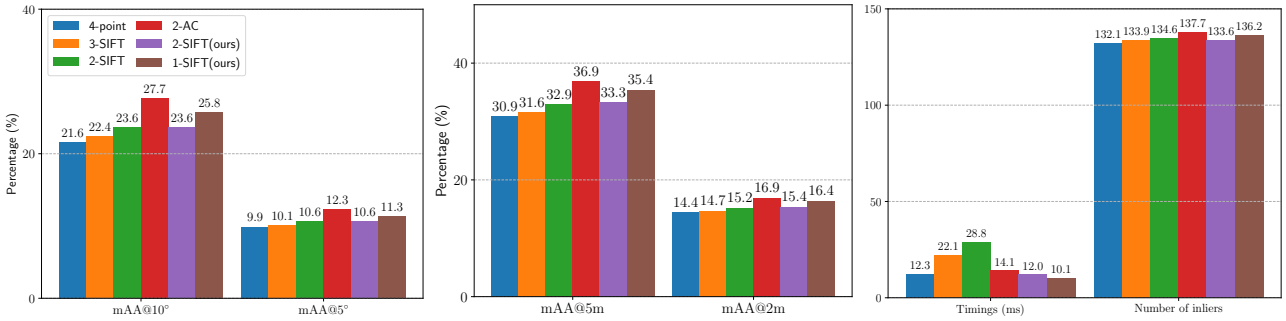
*ternational Conference on Computer Vision Theory and Application*. SciTePress, 2016. 2

[3] Daniel Barath and Zuzana Kukelova. Homography from two orientation-and scale-covariant features. In *CVPR*, pages 1091–1099, 2019. 2

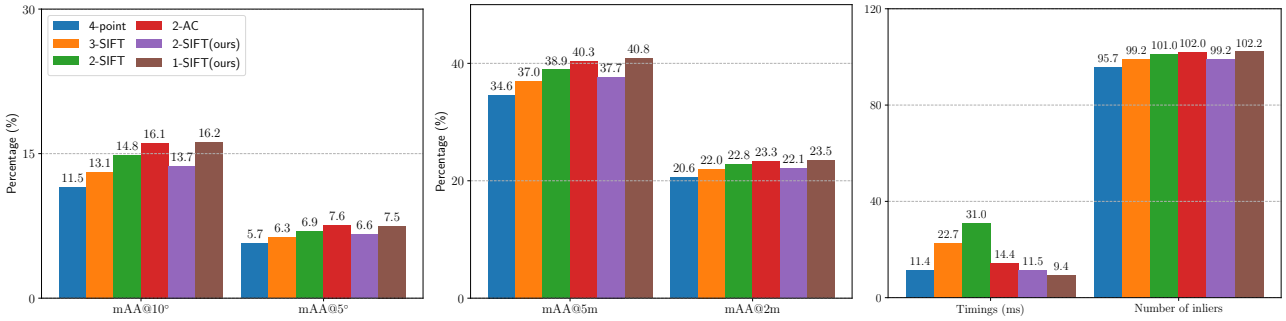
[4] Daniel Barath, Dmytro Mishkin, Michal Polic, Wolfgang Förstner, and Jiri Matas. A large-scale homography benchmark. In *CVPR*, pages 21360–21370, 2023. 3

[5] Matthew Brown, Richard I Hartley, and David Nistér. Minimal solutions for panoramic stitching. In *CVPR*, 2007. 1, 2

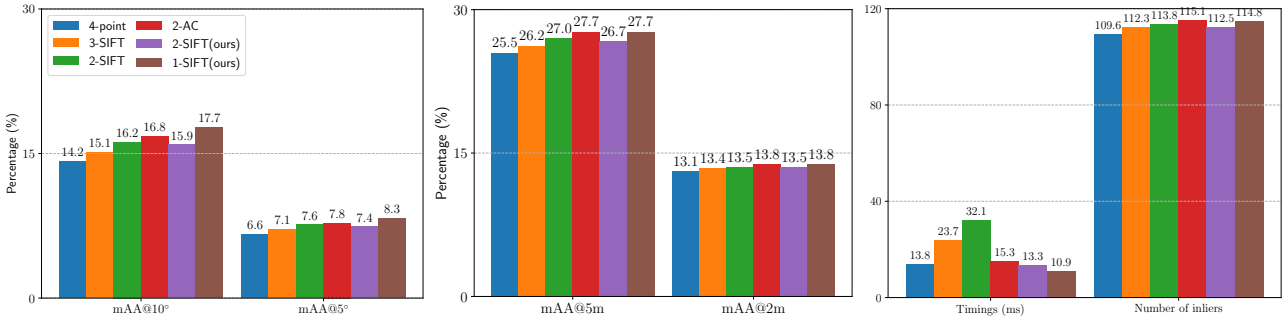
[6] Richard Hartley and Hongdong Li. An efficient hidden variable approach to minimal-case camera motion estimation. *IEEE TPAMI*, 34(12):2303–2314, 2012. 1



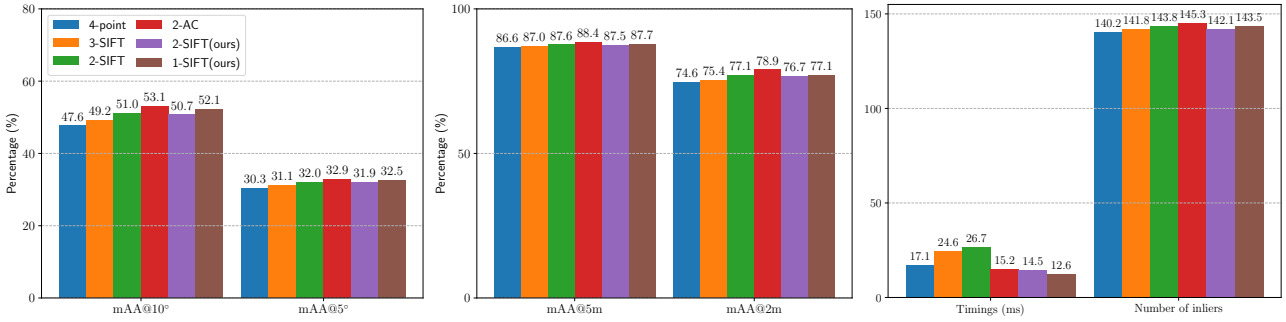

(a) Yorkminster




(b) Tower of London




(c) Madrid Metropolis



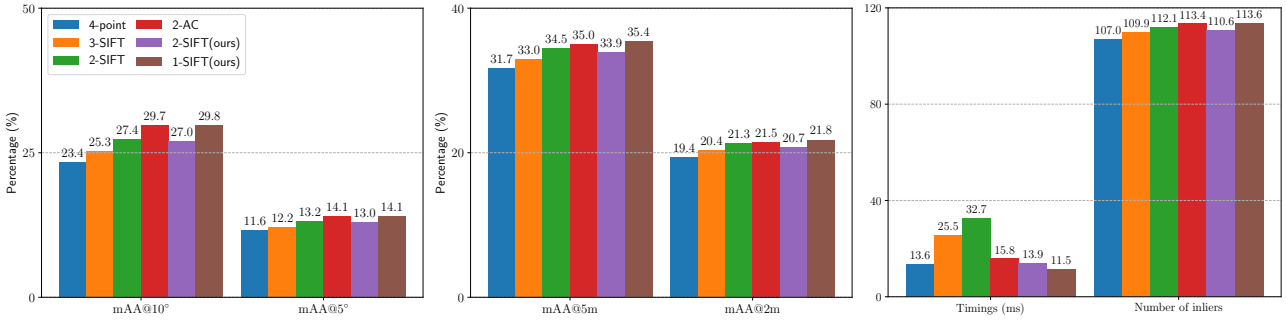

(d) Ellis Island

Figure 3. **Per-scene results on the HEB dataset.** (a) Yorkminster, (b) Tower of London, (c) Madrid Metropolis, and (d) Ellis Island.

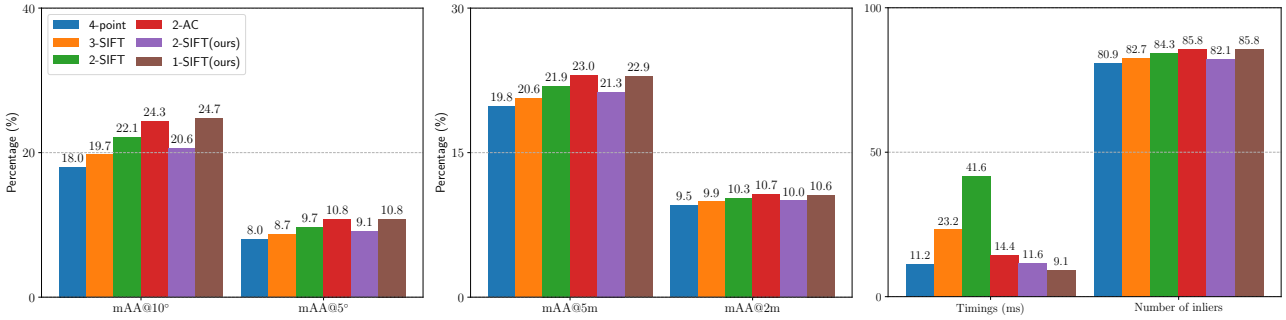

[7] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 1, 2, 3

[8] Berthold KP Horn. Closed-form solution of absolute orientation using unit quaternions. *Josa a*, 1987. 1, 2

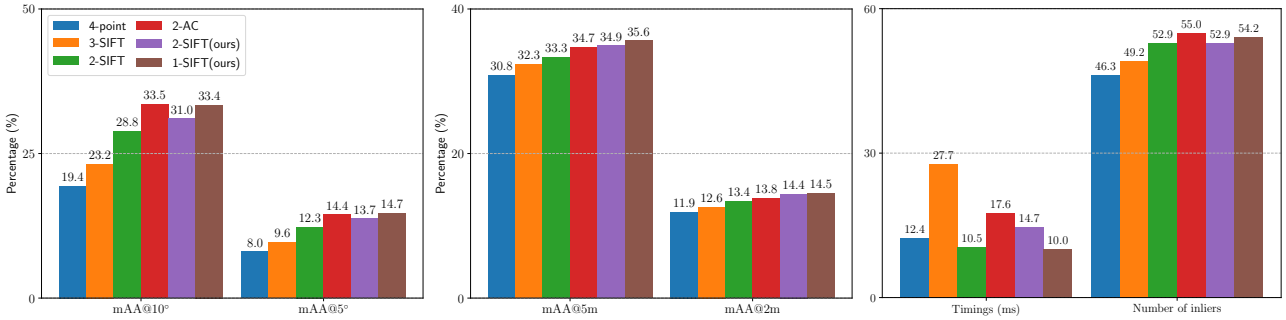
[9] Jongmin Lee, Yoonwoo Jeong, and Minsu Cho. Self-supervised learning of image scale and orientation. In *BMVC*, 2021. 2



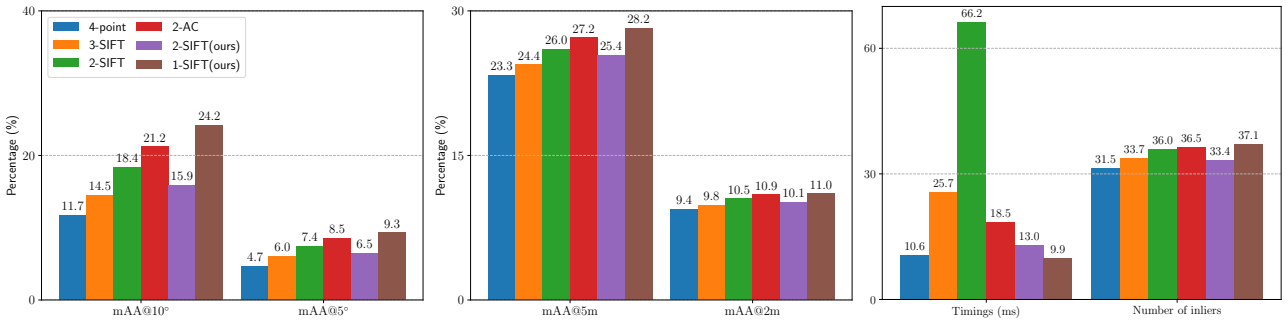

(a) Roman Forum




(b) Vienna Cathedral



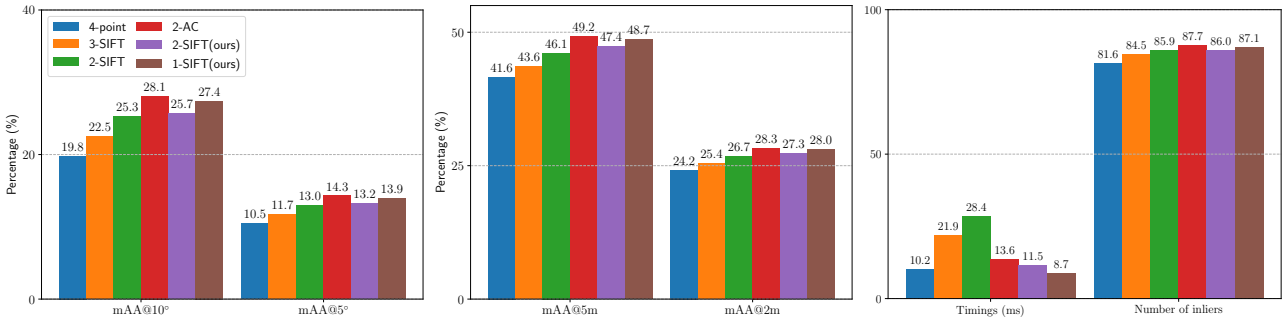

(c) Piazza del Popolo



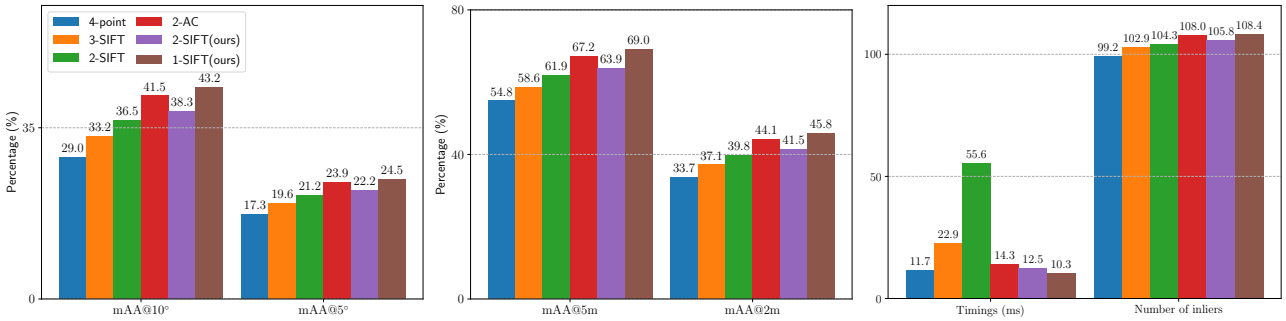

(d) Union Square

Figure 4. **Per-scene results of the HEB dataset.** (a) Roman Forum, (b) Vienna Cathedral, (c) Piazza del Popolo, and (d) Union Square.

[10] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 2




(a) NYC Library




(b) Alamo

Figure 5. **Per-scene results of the HEB dataset.** (a) NYC Library, (b) Alamo.