This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Noisy One-point Homographies are Surprisingly Good

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

Abstract

Two-view homography estimation is a classic and fundamental problem in computer vision. While conceptually simple, the problem quickly becomes challenging when multiple planes are visible in the image pair. Even with correct matches, each individual plane (homography) might have a very low number of inliers when comparing to the set of all correspondences. In practice, this requires a large number of RANSAC iterations to generate a good model hypothesis. The current state-of-the-art methods therefore seek to reduce the sample size, from four point correspondences originally, by including additional information such as keypoint orientation/angles or local affine information. In this work, we continue in this direction and propose a novel one-point solver that leverages different approximate constraints derived from the same auxiliary information. In experiments we obtain state-of-the-art results, with execution time speed-ups, on large benchmark datasets and show that it is more beneficial for the solver to be sample efficient compared to generating more accurate homographies.

1. Introduction

Scenes with dominant planes are common in man-made environments, and also provide strong cues on the 3D scene geometry. In these scenarios, finding the two-view camera geometry can be formulated as an image homography estimation task. Assuming the cameras' intrinsic calibrations are known, each homography can be directly factorized into the 3D plane parameters and the relative camera pose. This allows for a very compact representation of the entire geometric setup, using only a single 3×3 matrix without any additional internal constraints.

The standard approach for finding homographies relies on first detecting a sparse set of tentative matching keypoints, followed by robust estimation in some variant of RANdom SAmple Consensus (RANSAC) [25]. At the core of these frameworks is the process of repeatedly generating homography proposals from small subsets of the data, followed by verification and scoring using the full set of correspondences. To minimize the risk of including an outlier correspondence, it is preferred to estimate the homographies from as few correspondences as possible, so called *minimal samples*. For homography estimation, this is particularly important as "outliers" might come from correct matches that simply belong to a different plane (and thus do not support the sought homography), *i.e.* solving the model selection problem. So, even in favorable matching conditions, this can result in extremely low inlier ratios when considering all available correspondences between two images. In the recent large-scale Homography Estimation Benchmark (HEB) [14], this trend is clear where the top performing methods leverage additional information (e.g. orientation/angle or affine correspondences) to reduce the number of points necessary to generate the homography proposals inside RANSAC.

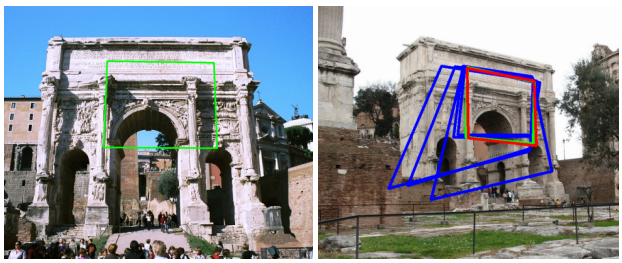


Figure 1. **One-point Homography Estimation.** A plane from the source image (green patch to the left) is transformed using homographies. Ground truth patch is shown in green in the target image to the right. Blue shows (noisy) transformed patches for the top five inlier samples using the proposed 1-SIFT solver. Red corresponds to the estimated homography resulting from GC-RANSAC [9] with the proposed 1-SIFT solver.

In this paper, we further delve into the trade-off between model accuracy and model complexity. We combine seemingly redundant (and dependent) additional constraints from auxiliary information to derive a homography solver that only requires *a single point*. In our experiments we show that this solver, while yielding very noisy homography estimates, improves over the current state-of-the-art due to being extremely sample efficient.

1.1. Related Work

Robust Estimators. The standard paradigm for robust model estimation is to use a hypothesize-and-verify framework such as RANSAC [25]. These methods generate model proposals by randomly sampling small subsets of the data and fitting models to these. Samples which only contain inliers are likely to yield good models which have a large consensus set (measurements that agree with the particular model). Since Fischler and Bolles' original paper [25], there have been many proposed improvements to the original algorithm, for example better sampling [19, 43, 45], improved scoring [10–12], early stopping criteria [39], and early sample rejection [17]. The overall research area is still active and new variants are published yearly [18, 43, 47].

In [20] Chum et al. highlighted the fact that even all-inlier samples might lead to poor models due to measurement noise or unstable configurations. To address this they proposed LO-RANSAC, which performs a local optimization, iteratively refitting the model estimates on the tentative inlier sets whenever promising candidates are found. Since then, there have been multiple proposed approaches on how to do this refitting best, e.g. LO$^+$-RANSAC [32], MAGSAC++ [10] and GC-RANSAC [9]. Since these methods are designed to be robust against low-quality model proposals, they are also well suited to handle modeling errors. In this paper, we leverage this robustness and propose new minimal estimators that yield very noisy homographies, due to relying on approximate geometric constraints.

Homography Estimators. At the core of RANSAC are the model estimators (minimal solvers) that, given a sample, generate one or more model proposals. For a general homography, four point correspondences is minimal and the homography can be estimated linearly [30]. In many image pairs, multiple planes are visible, which means that the effective inlier-ratio for each individual plane (homography) can be extremely low when considering the full set of matches. Note that these "outliers" occur even with perfect point-wise matching. To tackle this issue, there have been several methods proposed that leverage additional information or assumptions to reduce the number of points (matches) necessary to generate proposals. In [4], Barath and Hajder propose a homography solver that uses two affine correspondences (point coordinates together with a 2×2 local affine transformation). The affine correspondences can be estimated directly from image data [15, 40] or approximated using scale/orientation estimates from e.g. SIFT [37]. In [3, 7] Barath et al. proposed minimal estimators that instead directly use the SIFT scale and orientation to estimate the homography. There have also been papers that introduce other constraints, e.g. [5] propose using known epipolar geometry together with one affine correspondence, while [26] use a single affine aware correspondence to estimate the affine homography. In [13, 22, 24], the authors show that the relative orientation can be partially replaced with information about the gravity direction in both images. In [23] the authors combine SIFT and gravity for homography estimation. Auxiliary point-wise information has also been used to reduce the sample complexity in other geometric estimation tasks, e.g. essential matrix [1, 6, 8, 35], fundamental matrix [3], absolute pose [46], and generalized relative pose [27–29]. Another approach is to use approximate or simplified model assumptions. For example [44] leverages affine fundamental matrices and [42] consider orthographic approximations for essential matrix estimation.

2. Background

Assume that we have a set of 3D points $\{\mathbf{X}_i\}, i = 1, 2, 3, ..., k$ that lie on a plane in 3D space. These are observed by two cameras $\mathbf{K}_1[\mathbf{I} \mid \mathbf{0}]$ and $\mathbf{K}_2[\mathbf{R} \mid \mathbf{t}]$ with 2D-image point correspondences $\{\mathbf{m}_{i1}, \mathbf{m}_{i2}\}, i = 1, 2, 3, ..., k$ in the first and second images, respectively. We focus on the calibrated setting where $\mathbf{K}_1$ and $\mathbf{K}_2$ are known. In the supplementary material, we discuss the uncalibrated case. With known calibration we have $\mathbf{x}_{i2} = \mathbf{K}_2^{-1}\mathbf{m}_{i2}$, and $\mathbf{x}_{i1} = \mathbf{K}_1^{-1}\mathbf{m}_{i1}$. This gives

$$\alpha_{i1}\mathbf{x}_{i1} = \mathbf{X}_i, \tag{1}$$

$$\alpha_{i2}\mathbf{x}_{i2} = \mathbf{R}\mathbf{X}_i + \mathbf{t}, \tag{2}$$

where α_{i1} and α_{i2} are the depths of the points $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$, respectively. Let $\mathbf{n}$ be the unit normal vector of the plane with respect to the first camera frame, and let d denote the distance from the plane to the optical center of the first camera. Then we have

$$\mathbf{n}^\top \mathbf{X}_i = d. \tag{3}$$

Substituting (1) into (3) we get

$$\frac{1}{\alpha_{i1}} = \frac{1}{d}\mathbf{n}^\top \mathbf{x}_{i1}. \tag{4}$$

On the other hand, from (1) and (2) we have

$$\frac{\alpha_{i2}}{\alpha_{i1}}\mathbf{x}_{i2} = \mathbf{R}\mathbf{x}_{i1} + \frac{1}{\alpha_{i1}}\mathbf{t}. \tag{5}$$

Substituting (4) into the factor of $\mathbf{t}$ in (5), we get

$$\frac{\alpha_{i2}}{\alpha_{i1}}\mathbf{x}_{i2} = \underbrace{(\mathbf{R} + \frac{\mathbf{t}}{d}\mathbf{n}^\top)}_{\mathbf{H}}\mathbf{x}_{i1}, \tag{6}$$

where

$$\mathbf{H} = \mathbf{R} + \frac{\mathbf{t}}{d}\mathbf{n}^\top \tag{7}$$

is the Euclidean homography matrix.

	Methods	Reference	Number of solutions	Requires known intrinsics	Estimates focal length	Constraints			
						Point	Affine	Line	Depth
General	4-point	[30]	1	-	-	✔	-	-	-
	3-SIFT	[3]	1	-	-	✔	-	✔	-
	2-AC	[4]	1	-	-	✔	✔	-	-
	2-SIFT	[7]	4	-	-	✔	✔	✔	-
	2-SIFT	Proposed	2	✔	-	✔	-	✔	✔
	1-SIFT	Proposed	2	✔	-	✔	✔	✔	✔
Rotation	2-point(rot)	[31]	1	✔	-	✔	-	-	-
	2-point(f)	[16]	3	-	✔	✔	-	-	-
	1-SIFT(rot)	Proposed	1	✔	-	✔	-	✔	-
	1-SIFT(f)	Proposed	1	-	✔	✔	-	✔	✔

Table 1. **Overview of homography solvers.** The properties and constraints for the proposed (gray) and state-of-the-art solvers. Note that while we call some solvers 1/2/3-SIFT, the solvers can be used with any keypoint detector as long as scale and orientation information is available.

2.1. Orientation and Scale Constraints

Many widely-used feature detectors, e.g. SIFT and SURF, do not only provide point correspondences but also additional information about each feature's scale and rotation. We will now describe how such information directly can provide constraints on the homography.

Orientation Constraint. If we assume that the feature orientation can be considered as the direction of a line passing through this point, then a point correspondence with orientation can provide both point and line homography constraints. Such a correspondence will provide three linearly independent constraints for homography estimation.

Consider a line $\mathbf{l}_{i1}$ passing through the point $\mathbf{x}_{i1}$ in the first image, that maps to the line $\mathbf{l}_{i2}$ passing through the point $\mathbf{x}_{i2}$ in the second image, i.e.

$$\mathbf{l}_{i1}^\top \mathbf{x}_{i1} = 0, \tag{8}$$

$$\mathbf{l}_{i2}^\top \mathbf{x}_{i2} = 0. \tag{9}$$

Compared to points, lines map with the inverse of the transposed homography [30], i.e.

$$\lambda_i \mathbf{l}_{i2} = \mathbf{H}^{-\top} \mathbf{l}_{i1} \implies \lambda_i^{-1} \mathbf{l}_{i1} = \mathbf{H}^\top \mathbf{l}_{i2}. \tag{10}$$

From this we can eliminate the scale factor λ_i to get

$$[\mathbf{l}_{i1}]_\times \mathbf{H}^\top \mathbf{l}_{i2} = \mathbf{0}, \tag{11}$$

where $[\mathbf{l}_{i1}]_\times$ is the skew-symmetric matrix of $\mathbf{l}_{i1}$. In general, a pair of lines provides two constraints for homography estimation. However, as the lines are passing through the point correspondence (8)-(9), they only yield one additional linearly independent constraint.

Scale Constraint. In this paper, we assume that the relative depth can be approximated from relative SIFT [36] scale.

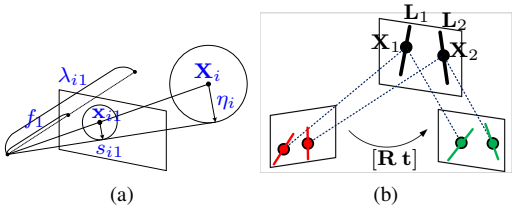


Figure 2. **Orientation and scale constraints.** (a) The SIFT scale in the image can be considered as an object with radius η and depth α in the 3D space projected into the image with focal length f and radius s. (b) Corresponding lines are mapped by the inverse of the transposed homography.

This assumption was also used in previous works [1, 27, 35]. Based on Fig. 2, we have

$$\frac{s_{i1}}{\eta_i} = \frac{f_1}{\alpha_{i1}} \text{ and } \frac{s_{i2}}{\eta_i} = \frac{f_2}{\alpha_{i2}}. \tag{12}$$

Hence

$$\sigma_i = \frac{\alpha_{i2}}{\alpha_{i1}} = \frac{f_2 s_{i1}}{f_1 s_{i2}}. \tag{13}$$

In this case, the relative depth σ_i can be formulated using the focal lengths (f_1, f_2) and the SIFT scales (s_{i1}, s_{i2}).

Let $\mathbf{H} = [h_1, h_2, h_3; h_4, h_5, h_6; h_7, h_8, h_9]$, $\mathbf{x}_{11} = [u_{11}\ v_{11}\ 1]^\top$ and $\mathbf{x}_{12} = [u_{12}\ v_{12}\ 1]^\top$. One point correspondence with known relative depth $\sigma_1 \mathbf{x}_{12} = \mathbf{H}\mathbf{x}_{11}$ then provides three constraints for homography estimation:

$$\begin{aligned} u_{11}h_1 + v_{11}h_2 + h_3 - \sigma_1 u_{12} &= 0, \\ u_{11}h_4 + v_{11}h_5 + h_6 - \sigma_1 v_{12} &= 0, \\ u_{11}h_7 + v_{11}h_8 + h_9 - \sigma_1 &= 0. \end{aligned} \tag{14}$$

Combining (11) and (14) we may obtain five constraints from one SIFT correspondence. However, since the line is passing through the point, only four of the constraints are linearly independent.

2.2. Affine Constraints

For an affine correspondence we have a triplet $([u_{11}, v_{11}], [u_{12}, v_{12}], \mathbf{A}_1)$, where $\mathbf{A}_1$ is a 2×2 matrix coding the local affine transformation. We denote the elements of $\mathbf{A}_1$ with (in row-major order) a_1, a_2, a_3, and a_4. To define $\mathbf{A}_1$, we use the formulation provided in [41], given as the first-order Taylor-approximation of the projection functions. For perspective cameras, the formula for $\mathbf{A}_1$ is the first-order approximation of the related homography matrix $\mathbf{H}$ given by

$$\begin{aligned} a_1 &= \frac{\partial u_{12}}{\partial u_{11}} = \frac{h_1 - u_{12}h_7}{\sigma_1}, \quad a_2 = \frac{\partial u_{12}}{\partial v_{11}} = \frac{h_2 - u_{12}h_8}{\sigma_1}, \\ a_3 &= \frac{\partial v_{12}}{\partial u_{11}} = \frac{h_4 - v_{12}h_7}{\sigma_1}, \quad a_4 = \frac{\partial v_{12}}{\partial v_{11}} = \frac{h_5 - v_{12}h_8}{\sigma_1}, \end{aligned} \tag{15}$$

where $\sigma_1 = u_{11}h_7 + v_{11}h_8 + h_9$ is the projective depth or relative depth. After re-arranging (15), four linear constraints are obtained from $\mathbf{A}_1$, namely

$$\begin{aligned} h_1 - u_{12}h_7 - a_1\sigma_1 &= 0, \\ h_2 - u_{12}h_8 - a_2\sigma_1 &= 0, \\ h_4 - v_{12}h_7 - a_3\sigma_1 &= 0, \\ h_5 - v_{12}h_8 - a_4\sigma_1 &= 0. \end{aligned} \tag{16}$$

In general, the local affine transformation hence provides four linear constraints for homography estimation.

Local Affine Transformation from Scale and Orientation. Given keypoint orientation (o_{ij}) and scale (s_{ij}) information, the local affine shape can be (poorly) approximated as

$$\mathbf{A}_{ij} = s_{ij} \begin{bmatrix} \cos(o_{ij}) & -\sin(o_{ij}) \\ \sin(o_{ij}) & \cos(o_{ij}) \end{bmatrix}. \tag{17}$$

The affine correspondence can then be approximated as

$$\mathbf{A}_i = \mathbf{A}_{i2}\mathbf{A}_{i1}^{-1}. \tag{18}$$

In this paper, we will discuss using different combinations of point, affine, line and depth constraints for homography estimation. Our focus is on general Euclidean homographies (i.e. calibrated and planar scene) and in the supplementary material we discuss the special case of pure rotation where a shared focal length can be estimated as well. A brief summary of all discussed solvers is shown in Table 1.

3. Euclidean Homography Estimation

We now present two minimal solvers for Euclidean homography estimation that leverage keypoint orientation and scale. First in Section 3.1, a two-point solver which uses the orientation and scale constraints, and next in Section 3.2, a one-point solver which also leverages the affine constraints.

3.1. The 2-SIFT Solver

Given two SIFT correspondences, we have

$$\sigma_1 \mathbf{x}_{12} = \mathbf{H}\mathbf{x}_{11}, \tag{19}$$
$$\sigma_2 \mathbf{x}_{22} = \mathbf{H}\mathbf{x}_{21}, \tag{20}$$
$$\mathbf{l}_{11} \sim \mathbf{H}^\top \mathbf{l}_{12}, \tag{21}$$
$$\mathbf{l}_{21} \sim \mathbf{H}^\top \mathbf{l}_{22}, \tag{22}$$
$$\mathbf{l}_{22} \times \mathbf{l}_{12} \sim \mathbf{H}(\mathbf{l}_{21} \times \mathbf{l}_{11}), \tag{23}$$

where $\sim$ denotes equality up to scale. Here (21)-(22) is linearly dependent on (19),(20),(23). In fact, (23) is simply the point-match coming from intersecting the two lines. We can write these equations in matrix form

$$\mathbf{Z} = \mathbf{H}\mathbf{Y}, \tag{24}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are 3×3 matrices that only depend on image data, except for an unknown scale factor λ

$$\begin{aligned} \mathbf{Y} &= [\mathbf{x}_{11}, \mathbf{x}_{21}, \mathbf{l}_{21} \times \mathbf{l}_{11}], \\ \mathbf{Z} &= [\sigma_1 \mathbf{x}_{12}, \sigma_2 \mathbf{x}_{22}, \lambda(\mathbf{l}_{22} \times \mathbf{l}_{12})]. \end{aligned} \tag{25}$$

This directly gives an expression for the homography matrix

$$\mathbf{H} = \mathbf{Z}\mathbf{Y}^{-1}. \tag{26}$$

As shown in [38, 48], a Euclidean homography matrix should satisfy the singular value constraint

$$\text{median}(\text{svd}(\mathbf{H})) = 1, \tag{27}$$

where the second largest singular value of $\mathbf{H}$ should be 1. Hence $\mathbf{H}^\top\mathbf{H}$ should have a unit eigenvalue, and we have

$$\det(\mathbf{H}^\top\mathbf{H} - \mathbf{I}_3) = 0. \tag{28}$$

Naively this seems to yield a degree six equation in λ as $\mathbf{H}$ is linear in λ. However, due to the structure of $\mathbf{H}$ it turns out that (28) is in fact only quadratic in λ. There are up to two solutions that ensure $\mathbf{H}$ has one unit singular value.

To see this, consider $\det(\mathbf{H}^\top\mathbf{H}) = (\det(\mathbf{H}))^2$. Based on (26), we have $\det(\mathbf{H}) = \det(\mathbf{Z})\det(\mathbf{Y}^{-1})$. Note that, $\det(\mathbf{Y}^{-1})$ is a constant, and $\det(\mathbf{Z})$ is linear in λ. In this case, $\det(\mathbf{H}^\top\mathbf{H})$ is a quadratic equation in λ. The degree of λ in the remaining parts of (28) would not be larger than two. Hence, (28) is a quadratic equation in λ.

3.2. The 1-SIFT Solver

We now present a minimal solver that estimates a homography from a single correspondence with associated scale and orientation. The idea is to leverage the relative depth (14) and line (11) constraints (as in the previous section), together with the affine correspondence constraints (15), where the AC is approximated using (17).

Given one SIFT correspondence, we have eight inhomogeneous linear constraints on the elements of the unknown calibrated homography $\mathbf{H}$, if we combine the affine constraints with the orientation and scale information. However, only seven of the eight constraints are linearly independent. To make a full-rank system, we use one of the constraints from mapping the line normals as if they were points with the homography. In the supplementary material, we show that surprisingly this *incorrect* line constraint has only a very minor impact on the results. Note that using this formulation we can not fix the scale of the homography arbitrarily. However, we can write these equations as

$$\mathbf{B}\hat{\mathbf{h}} = \mathbf{0}, \tag{29}$$

where $\hat{\mathbf{h}} = [h_1, h_2, ..., h_9, 1]^\top$ are the nine entries of $\mathbf{H}$, extended with a one, and $\mathbf{B}$ is a matrix of size 8×10 with known entries. The vector $\hat{\mathbf{h}}$ can be written as a linear combination of the two basis vectors from the two-dimensional null space of the matrix $\mathbf{B}$ as

$$\hat{\mathbf{h}} = \lambda_1 \hat{\mathbf{h}}_1 + \lambda_2 \hat{\mathbf{h}}_2, \tag{30}$$

with

$$\lambda_1 \hat{\mathbf{h}}_1(10) + \lambda_2 \hat{\mathbf{h}}_2(10) = 1, \tag{31}$$

where $\hat{\mathbf{h}}_1(10), \hat{\mathbf{h}}_2(10)$ are the tenth elements. In this case, the Euclidean homography matrix $\mathbf{H}$ can be formulated with a single unknown λ_1 (or λ_2). Substituting this formulation into the singular value constraints (27), we obtain a quadratic equation in λ_1, thus yielding two possible solutions for the homography matrix.

In contrast to classical homography estimation, the two proposed solvers require known intrinsic calibration, which is indeed a limitation of the method. However, in many practical scenarios a rough focal length is often known (*e.g.*, from EXIF) which is usually enough. In our experiments we show that even very coarse estimates of the focal length and principal point is sufficient.

Why Does This Work? The proposed solver use the same measurement keypoint scale twice, both in the relative depth and in the affine constraints (Sec. 2.2). While these constraint then seem to be redundant (as they come from the same measurements), they are algebraically independent as they rely on different approximations (using relative keypoint scale as relative depth, compared to the AC modeling the point-wise linearization of plane-induced homography). From a conceptual point-of-view, using these seemingly redundant measurements can be thought of as a way of randomly drawing homographies that are somewhat reasonable (at least satisfying the point correspondence, and being roughly correct in orientation/scale/etc.). However, the dependency between the input data in the two constraint sets means we get some very unstable estimates where part of the noise comes from this modeling error. But, as our experiments will show (Sec. 4), this can be sufficient to find initial inlier sets that allow the local optimization to converge to good homographies.

4. Experiments

To evaluate the proposed minimal estimators we use the recently proposed Homography Estimation Benchmark (`HEB`) [14][1] which is a large-scale homography benchmark, consisting of ten scenes that contain 226 260 homographies and includes roughly 4M correspondences. The dataset contains many image pairs that undergo significant viewpoint and illumination changes, leading to some instances with extremely low inlier ratios. For the experiments, we use the RootSIFT correspondences provided by the dataset.

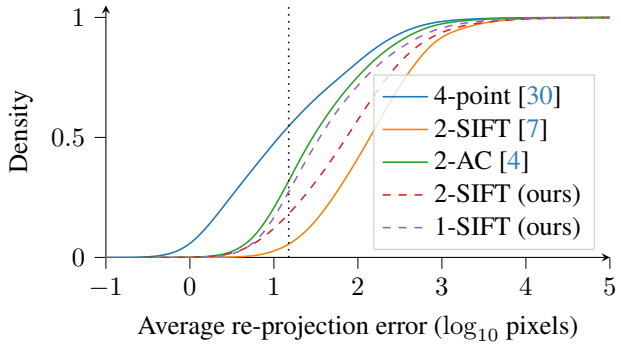
We also consider image pairs from the `HPatches` [2] dataset. This dataset is split into pairs that contain either viewpoint or illumination changes. In the supplementary material we show additional results, including a self-captured dataset exhibiting pure rotational motion.

4.1. Solver Stability

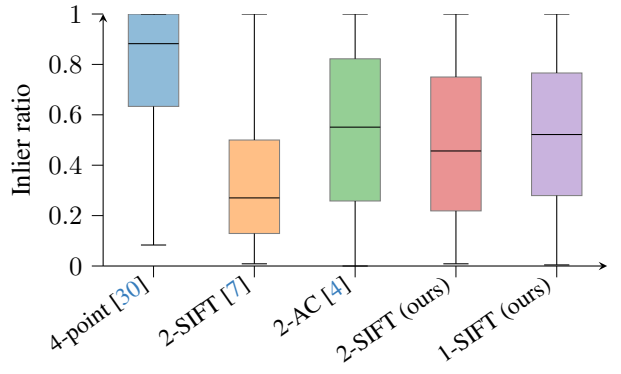
We first evaluate the characteristics of the proposed solvers in isolation. Since it is difficult to generate realistic noise for keypoint positions, orientations, and scales, we instead opt to perform psuedo-synthetic experiments using real image pairs. We consider the *Piazza del Popolo* scene from the `HEB` dataset. For each image pair, we extract the set of ground-truth inlier correspondences. For each solver, we then randomly draw as many all-inlier minimal samples as there are point-correspondences. Note that for the one-point solvers, this corresponds to exhaustive sampling, while for the other solvers we only explore a subset of the possible samples, similarly to what happens in RANSAC. For each minimal sample, we then calculate the re-projection error of all ground-truth matches with respect to the proposed homography. Fig. 3 (a) shows the cumulative distribution functions (CDFs) of the average re-projection errors across all ground truth matches. Fig. 3 (b) shows boxplots of the ratio of ground-truth inliers that were also inliers to the generated homographies. As expected, the 4-point solver yields the most stable estimates when applied to inlier correspondences, while the solvers that integrate auxiliary information are more noisy. Furthermore, we can see that the proposed 1-SIFT solver has similar accuracy to the 2-AC solver from [4], and that the proposed 2-SIFT solver is more accurate compared to the 2-SIFT solver from [7]. As the proposed 1-SIFT solver is outperforming the proposed 2-SIFT solver, we will focus our attention to the one point solver in the following experiments.

Note also that, while the above experiments show that the 4-point solver is significantly more stable on all-inlier sam-

[1] https://github.com/danini/homography-benchmark



(a)



(b)

Figure 3. **Solver stability.** (a) The cumulative distribution functions (CDF) of the re-projection errors on ground-truth inliers, using different minimal solvers on the scene "Piazza del Popolo" in the HEB dataset [14]. The dotted vertical line illustrates inlier threshold. (b) Distribution of inlier ratios for each estimated homography in the evaluation of (a).

ples compared to the other solvers, when running RANSAC it is enough to find a single model with sufficient inlier support from which local optimization can converge.

4.2. Qualitative Evaluation

To gain a better understanding of how our solver performs on minimal samples, we show some qualitative examples of the resulting homographies. We do this by again sampling from the ground-truth inlier matches. Again, we draw as many minimal samples as we have inlier matches, resulting in exhaustive sampling for the 1-point solver and random sampling otherwise. We compute the resulting homographies and sort the solutions based on number of inliers (w.r.t. error less than 20 pixels). In Fig. 4, we visualize the best homographies found by the 4-point solver and our 1-SIFT solver, for some image pairs in the `HEB` dataset. We also show the output from running the solvers in Graph-Cut RANSAC[2] [9] (GC-RANSAC). In the figure, we see that the homographies estimated from the 1-point solver are much more noisy than those estimated by the 4-point solver. In fact, in some cases even the best sampled homography is quite far from the ground truth (in green). However, as we see by the GC-RANSAC estimations (in red), even these noisy estimates can be enough for the local optimization and resampling in GC-RANSAC to find the correct homography. Note that the visualized minimal samples are drawn from inlier correspondences, and might not be sampled in an actual RANSAC. Indeed, in some of the examples, the 4-point GC-RANSAC fails to find any good models. In the supplementary, we show more qualitative examples.

[2] https://github.com/danini/graph-cut-ransac

4.3. Evaluation in Robust Estimation

We now evaluate the proposed solvers in the context of robust estimation. For the evaluation, we again integrate the solvers into GC-RANSAC, as it is representative of the state-of-the-art and since it has empirically been shown to be very robust against noisy model estimates. In GC-RANSAC (and other locally optimized RANSACs), two different solvers are used: (a) one for estimating the pose from a minimal sample and (b) one for fitting to a larger-than-minimal sample when doing final pose polishing on all inliers or in the local optimization step. For (a), the main objective is to solve the problem using as few correspondences as possible since the processing time depends exponentially on the number of correspondences required for the pose estimation. We tested six minimal solvers, the proposed 1-SIFT and 2-SIFT solvers, the standard 4-point solver [30], the existing state-of-the-art 2-SIFT solver [7], the 3-SIFT solver [3], and the 2-AC solver [4]. The purpose of (b) is to estimate the pose parameters as accurately as possible. For (b), we used the 4-point solver to do the non-minimal re-fitting. For the experiments we use the two discussed datasets `HEB` and `HPatches`.

Comparison to SOTA. Table 2 reports the mean Average Accuracy (mAA) of rotation errors at 5° and 10°, mAA of absolute translation errors at 2 m and 5 m, average processing times (in ms), and average number of inliers on the `HEB` dataset. The proposed 1-SIFT solver achieves comparable accuracy to the state-of-the-art 2-AC solver [4], while having significantly lower runtime ($\approx 30\%$ lower on `HEB`) due to the reduced sample size requiring fewer iterations. This is further illustrated in Figure 5 which shows the runtime-accuracy trade-off. More detailed results are shown in the supplementary material.

Ablation Study of 1-SIFT. As an ablation study we experiment with different variations of the proposed 1-SIFT solver as well as different approaches for obtaining the keypoint scale and orientation. For the ablation study we use the `HPatches` dataset. As the dataset does not provide any intrinsic calibration, we simply set the principal point to the be at the center of the image, and the focal length to be $f = \max(width, height)$. The results are summarized in

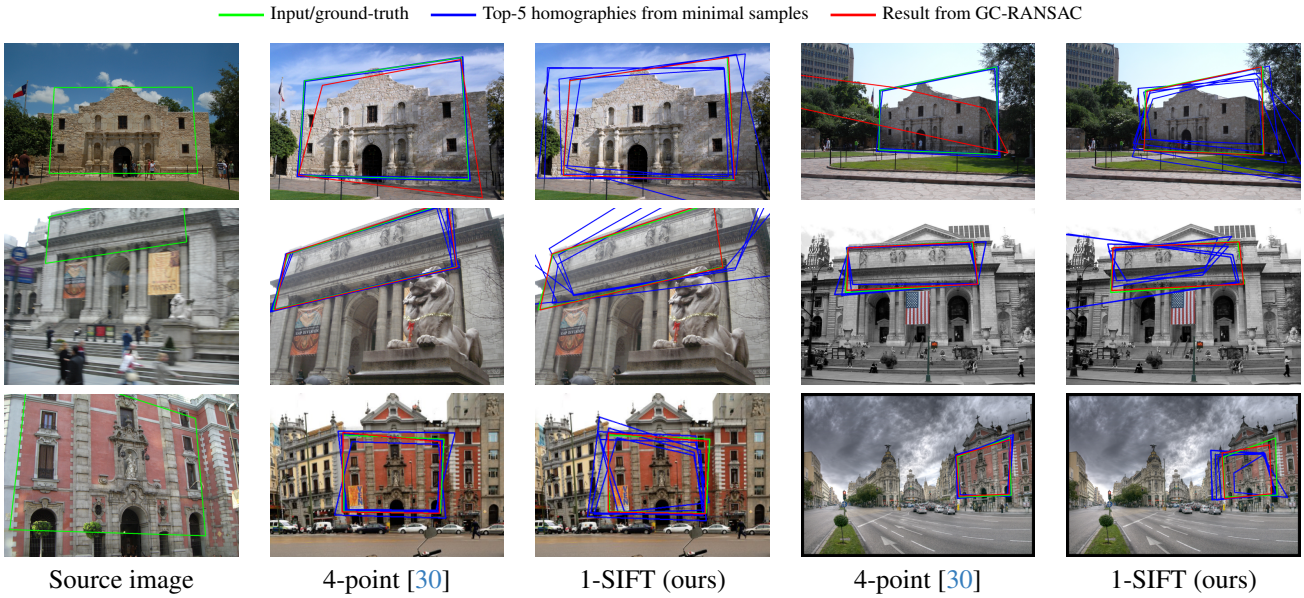


Figure 4. **Qualitative evaluation of minimal solver.** A plane from the source image is transformed to two other viewpoints by different homographies. The homographies are estimated using the traditional 4-point solver and our 1-SIFT solver. Blue quadrilaterals show the top-5 best homography estimates for each solver, on minimal samples drawn from ground-truth inliers. Red quadrilaterals show estimates obtained by running each solver in GC-RANSAC, on all correspondences. Ground-truth is 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 due to the low probability of finding a good minimal sample.

Solver	mAA@5°	mAA@10°	mAA@2m	mAA@5m	Nbr. Inl.	Runtime (ms)
4-point [30]	11.3	21.6	23.1	37.9	92.4	<u>12.4</u>
3-SIFT [3]	12.2	23.8	24.1	39.4	95.0	24.0
2-SIFT [7]	13.4	26.4	25.0	41.0	96.8	52.1
2-AC [4]	<u>14.6</u>	<u>29.1</u>	**26.2**	<u>43.0</u>	**98.6**	15.3
2-SIFT (ours)	13.4	26.2	25.2	41.2	95.8	12.8
1-SIFT (ours)	**14.7**	**29.4**	**26.2**	**43.1**	<u>98.3</u>	**10.3**

Table 2. **Evaluation on `HEB`.** The table shows the results on the `HEB` dataset. For each metric, we have highlighted the **best** and <u>second best</u> result. The proposed 1-SIFT solver provides similar accuracy as the 2-AC solver with significantly faster runtime.

Table 3; we report results separately for the illumination and viewpoint pairs, as well as the aggregate results on the full dataset. The table shows the AUC for the corner-warp error for different pixel thresholds. Where applicable, we also show the result from the `HEB` dataset.

We first compare the 1-SIFT solver with other possible 1-point approaches, such as the solver from [26], which used a single affine aware correspondence. We also experiment with creating a 2D similarity transform based on the relative scale and orientation (denoted H-Similarity in the table), which performs well on `HPatches` but worse on `HEB`.

We also explore different approaches for generating the keypoint scale and orientation, including randomly drawing them from uniform distributions, having a fixed unit scale and zero-orientation, and estimating them with Self-Sca-Ori [33]. For the illumination-varying image pairs, there is no viewpoint change and the ground-truth homography is simply the identity transform, which explains why the fixed orientation and scale performs so well.

Finally we compare using keypoint matches established using the state-of-the-art matcher LightGlue [34] with SuperPoint keypoints [21]. To obtain scales and orientation, we again use Self-Sca-Ori [33].

Improving Features with AffNet [40]. We also experiment with improving the SIFT-features using AffNet [40] by first converting the SIFT features from the `HEB` dataset to local affine frames (LAFs) and extract corresponding patches in the original images. Then, we use AffNet[3] on the patches to estimate new LAF shapes. The original LAF shape is replaced with the refined one. We consider two cases: i) Extract scale and orientation from the refined

[3] https://github.com/ducha-aiki/affnet

Solver	Keyp.	Scale	Orientation	Viewpoint 1px	Viewpoint 3px	Viewpoint 5px	Illumination 1px	Illumination 3px	Illumination 5px	Full 1px	Full 3px	Full 5px	HEB mAA@10°
1-SIFT (ours)	SIFT	SIFT	SIFT	**25.3**	**53.1**	**63.5**	32.9	62.3	72.0	28.8	56.8	**67.3**	**29.4**
1-SIFT (ours)	SIFT	1.0	0.0	22.7	46.6	56.1	**45.7**	**70.6**	**79.3**	**33.1**	**57.9**	67.1	27.8
1-SIFT (ours)	SIFT	$\mathcal{U}(0.5, 1.5)$	$\mathcal{U}(0, 2\pi)$	7.24	11.7	13.5	5.08	7.39	8.13	4.41	7.50	8.55	22.1
1-SIFT (ours)	SIFT	$\mathcal{U}(0.5, 1.5)$	SIFT	25.1	51.7	61.3	24.3	50.7	59.6	25.1	51.4	60.3	27.7
1-SIFT (ours)	SIFT	SIFT	$\mathcal{U}(0, 2\pi)$	6.21	11.5	13.1	5.49	8.34	9.15	4.64	8.14	9.26	22.0
1-SIFT (ours)	SIFT	SSO [33]	SSO [33]	24.9	51.9	62.2	30.2	61.7	71.3	26.8	56.0	66.4	-
1-SIFT (ours)	SIFT	SIFT	SIFT	25.3	53.1	63.5	32.9	62.3	72.0	28.8	56.8	67.3	**29.4**
HSolo [26]	SIFT	SIFT	SIFT	17.2	28.5	31.8	33.1	63.4	73.0	21.5	43.4	51.0	17.7
H-Similarity	SIFT	SIFT	SIFT	25.0	52.9	63.6	35.1	62.7	71.7	28.4	56.4	66.9	27.8
4-point [30]	SIFT	-	-	**29.4**	**56.6**	**67.7**	**38.0**	**65.4**	**75.2**	**30.0**	**60.1**	**71.3**	21.6
1-SIFT (ours)	SP+LG	SSO [33]	SSO [33]	30.8	62.2	72.4	64.8	77.4	83.3	35.2	66.7	76.8	-
4-point [30]	SP+LG	-	-	**31.8**	**63.3**	**73.6**	**66.2**	**81.1**	**87.9**	**37.7**	**69.4**	**79.3**	-

Table 3. **Evaluation and ablation on `HPatches`.** The table shows the results on the `HPatches` dataset for different input combinations. The first six rows show an ablation on the input to our 1-SIFT solver. The final two rows show a comparison between our solver and the 4-point solver on SuperPoint+LightGlue (SP+SG) keypoints, using Self-Sca-Ori (SSO) for scale and orientation. In the last column, we also include a comparison with `HEB`.

LAFs, and combine them with the refined LAFs for 1-SIFT homography estimation. ii) Extract scale and orientation from the refined LAFs, and then convert the scale and orientation to new LAFs for 1-SIFT homography estimation. Table 4 shows that refining the affine shape with AffNet can slightly improve the estimation results.

Metric	SIFT	AffNet	AffNet→SIFT
mAA(R)@10°	33.4	**37.3**	36.6
mAA(R)@5°	14.7	**16.5**	**16.5**
mAA(T)@5m	35.6	**36.7**	36.3
mAA(T)@2m	14.5	**15.4**	14.5
Inliers	54.5	**56.3**	56.1
Timings(ms)	**8.5**	9.2	9.2

Table 4. **Results with improved features.** We convert the SIFT features to LAFs and apply AffNet to obtain refined LAFs. We compare original SIFT features with AffNet refined LAFs, and with converting the refined LAFs back to SIFT features.

5. Conclusion

In this paper we have proposed new solvers for estimating homographies from keypoints with associated scale and orientation. The solvers are extremely sample efficient, even being able to estimate a full homography from a single correspondence. This is achieved by leveraging the same input measurements (scale / orientation) in multiple different approximate geometric constraints. As expected, this yields very noisy homographies.

Surprisingly, our experiments show that SOTA robust estimations (such as GC-RANSAC [9]), which are designed to be robust to poor initial models, can offset the modeling noise introduced by using the seemingly dependent constraints. In the paper we have explored this trade-off between sample size and model noise, and show that for homography estimation in challenging conditions, it is more important to sample fewer points compared to generating more accurate initial models.

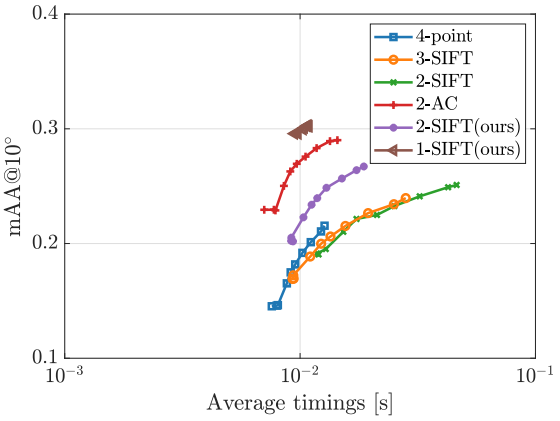


Figure 5. **Accuracy-runtime trade-off on `HEB`.** The plot show the trade-off between accuracy (mAA@10°) and runtime (seconds) for the `HEB` dataset. The plots were obtained by varying the maximum number of iterations for each method from 10 to 10^3.

Limitations. The proposed solver relies on each keypoint having an associated scale and orientation, making it more difficult to use state-of-the-art learned detectors which often only estimate the keypoint position. However, there are some recent works (e.g. Self-Sca-Ori [33]), which can predict scale and orientation for arbitrary keypoints. While our experiments show that the estimates obtained from [33] are still rather coarse, this might be improved in future works.

Acknowledgments. This work was supported by the strategic research project ELLIIT, the Swedish Research Council (grant no. 2023-05424), and the Czech Science Foundation (GACR) JUNIOR STAR Grant No. 22-23183M.

References

[1] Jonathan Astermark, Yaqing Ding, Viktor Larsson, and Anders Heyden. Fast relative pose estimation using relative depth. In *3DV*, 2024. 2, 3

[2] Vassileios Balntas, Karel Lenc, Andrea Vedaldi, and Krystian Mikolajczyk. Hpatches: A benchmark and evaluation of handcrafted and learned local descriptors. In *CVPR*, 2017. 5

[3] Daniel Barath. Five-point fundamental matrix estimation for uncalibrated cameras. In *CVPR*, pages 235–243, 2018. 2, 3, 6, 7

[4] Daniel Barath and Levente Hajder. Novel ways to estimate homography from local affine transformations. In *11th International Conference on Computer Vision Theory and Application*. SciTePress, 2016. 2, 3, 5, 6, 7

[5] Daniel Barath and Levente Hajder. A theory of point-wise homography estimation. *Pattern Recognition Letters*, 94:7–14, 2017. 2

[6] Daniel Barath and Levente Hajder. Efficient recovery of essential matrix from two affine correspondences. *IEEE TIP*, 27(11):5328–5337, 2018. 2

[7] Daniel Barath and Zuzana Kukelova. Homography from two orientation-and scale-covariant features. In *CVPR*, pages 1091–1099, 2019. 2, 3, 5, 6, 7

[8] Daniel Barath and Zuzana Kukelova. Relative pose from sift features. In *ECCV*, pages 454–469. Springer, 2022. 2

[9] Daniel Barath and Jiří Matas. Graph-cut RANSAC. In *CVPR*, 2018. 1, 2, 6, 8

[10] Daniel Barath, Jana Noskova, Maksym Ivashechkin, and Jiri Matas. Magsac++, a fast, reliable and accurate robust estimator. In *CVPR*, 2020. 2

[11] Daniel Barath, Jana Noskova, and Jiri Matas. Marginalizing sample consensus. *IEEE TPAMI*, 2021.

[12] Daniel Barath, Luca Cavalli, and Marc Pollefeys. Learning to find good models in ransac. In *CVPR*, 2022. 2

[13] Daniel Barath, Dmytro Mishkin, Luca Cavalli, Paul-Edouard Sarlin, Petr Hruby, and Marc Pollefeys. Affineglue: Joint matching and robust estimation. *arXiv preprint arXiv:2307.15381*, 2023. 2

[14] Daniel Barath, Dmytro Mishkin, Michal Polic, Wolfgang Förstner, and Jiri Matas. A large-scale homography benchmark. In *CVPR*, pages 21360–21370, 2023. 1, 5, 6

[15] Adam Baumberg. Reliable feature matching across widely separated views. In *CVPR*, 2000. 2

[16] Matthew Brown, Richard I Hartley, and David Nistér. Minimal solutions for panoramic stitching. In *CVPR*, 2007. 3

[17] Luca Cavalli, Marc Pollefeys, and Daniel Barath. Nefsac: Neurally filtered minimal samples. In *ECCV*, 2022. 2

[18] Luca Cavalli, Daniel Barath, Marc Pollefeys, and Viktor Larsson. Consensus-adaptive ransac. *arXiv preprint arXiv:2307.14030*, 2023. 2

[19] Ondrej Chum and Jiri Matas. Matching with prosac-progressive sample consensus. In *CVPR*, 2005. 2

[20] 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

[21] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *CVPRW*, 2018. 7

[22] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. An efficient solution to the homography-based relative pose problem with a common reference direction. In *ICCV*, 2019. 2

[23] Yaqing Ding, Daniel Barath, and Zuzana Kukelova. Homography-based egomotion estimation using gravity and sift features. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 2

[24] Yaqing Ding, Jian Yang, Jean Ponce, and Hui Kong. Homography-based minimal-case relative pose estimation with known gravity direction. *IEEE TPAMI*, 2020. 2

[25] 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*, 1981. 1, 2

[26] Antonio Gonzales, Cara Monical, and Tony Perkins. Hsolo: Homography from a single affine aware correspondence. *arXiv preprint arXiv:2009.05004*, 2020. 2, 7, 8

[27] Banglei Guan and Ji Zhao. Relative pose estimation for multi-camera systems from point correspondences with scale ratio. In *ACM MM*, pages 5036–5044, 2022. 2, 3

[28] Banglei Guan, Ji Zhao, Zhang Li, Fang Sun, and Friedrich Fraundorfer. Minimal solutions for relative pose with a single affine correspondence. In *CVPR*, pages 1929–1938, 2020.

[29] Banglei Guan, Ji Zhao, Daniel Barath, and Friedrich Fraundorfer. Minimal solvers for relative pose estimation of multi-camera systems using affine correspondences. *International Journal of Computer Vision*, 2023. 2

[30] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2, 3, 6, 7, 8

[31] Berthold KP Horn. Closed-form solution of absolute orientation using unit quaternions. *Josa a*, 1987. 3

[32] Karel Lebeda, Jirı Matas, and Ondrej Chum. Fixing the locally optimized RANSAC–full experimental evaluation. In *BMVC*, 2012. 2

[33] Jongmin Lee, Yoonwoo Jeong, and Minsu Cho. Self-supervised learning of image scale and orientation. In *BMVC*, 2021. 7, 8

[34] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. Lightglue: Local feature matching at light speed. In *ICCV*, 2023. 7

[35] Stephan Liwicki and Christopher Zach. Scale exploiting minimal solvers for relative pose with calibrated cameras. In *BMVC*, 2017. 2, 3

[36] D. G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999. 3

[37] David G Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 2

[38] Ezio Malis and Manuel Vargas Villanueva. Deeper understanding of the homography decomposition for vision-based control. *INRIA, Tech. Rep.*, 2007. 4

[39] Jiri Matas and Ondrej Chum. Randomized ransac with sequential probability ratio test. In *ICCV*, 2005. 2

[40] Dmytro Mishkin, Filip Radenovic, and Jiri Matas. Repeatability is not enough: Learning affine regions via discriminability. In *ECCV*, 2018. 2, 7

[41] József Molnár and Dmitry Chetverikov. Quadratic transformation for planar mapping of implicit surfaces. *Journal of mathematical imaging and vision*, 48:176–184, 2014. 4

[42] Magnus Oskarsson. Two-view orthographic epipolar geometry: Minimal and optimal solvers. *Journal of Mathematical Imaging and Vision (JMIV)*, 2018. 2

[43] Valter Piedade and Pedro Miraldo. Bansac: A dynamic bayesian network for adaptive sample consensus. In *ICCV*, 2023. 2

[44] James Pritts, Ondřej Chum, and Jiří Matas. Approximate models for fast and accurate epipolar geometry estimation. In *International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pages 106–111. IEEE, 2013. 2

[45] Philip Hilaire Torr, Slawomir J Nasuto, and John Mark Bishop. Napsac: High noise, high dimensional robust estimation-it's in the bag. In *BMVC*, 2002. 2

[46] Jonathan Ventura, Zuzana Kukelova, Torsten Sattler, and Dániel Baráth. P1ac: Revisiting absolute pose from a single affine correspondence. In *ICCV*, 2023. 2

[47] Tong Wei, Yash Patel, Alexander Shekhovtsov, Jiri Matas, and Daniel Barath. Generalized differentiable ransac. In *ICCV*, 2023. 2

[48] Zhongfei Zhang and Allen R Hanson. Scaled euclidean 3d reconstruction based on externally uncalibrated cameras. In *Proceedings of International Symposium on Computer Vision-ISCV*, pages 37–42. IEEE, 1995. 4