Making Affine Correspondences Work in Camera Geometry Computation

Daniel Barath[1,2], Michal Polic[3], Wolfgang Förstner[5], Torsten Sattler[3,4], Tomas Pajdla[3], and Zuzana Kukelova[2]

[1] Machine Perception Research Laboratory, SZTAKI in Budapest
[2] VRG, Faculty of Electrical Engineering, Czech Technical University in Prague
[3] Czech Institute of Informatics, Robotics and Cybernetics, CTU in Prague
[4] Chalmers University of Technology in Gothenburg
[5] Institute of Geodesy and Geoinformation, University of Bonn

Abstract. Local features *e.g.* SIFT and its affine and learned variants provide region-to-region rather than point-to-point correspondences. This has recently been exploited to create new minimal solvers for classical problems such as homography, essential and fundamental matrix estimation. The main advantage of such solvers is that their sample size is smaller, *e.g.*, only two instead of four matches are required to estimate a homography. Works proposing such solvers often claim a significant improvement in run-time thanks to fewer RANSAC iterations. We show that this argument is not valid in practice if the solvers are used naively. To overcome this, we propose guidelines for effective use of region-to-region matches in the course of a full model estimation pipeline. We propose a method for refining the local feature geometries by symmetric intensity-based matching, combine uncertainty propagation inside RANSAC with preemptive model verification, show a general scheme for computing uncertainty of minimal solvers results, and adapt the sample cheirality check for homography estimation. Our experiments show that affine solvers can achieve accuracy comparable to point-based solvers at faster run-times when following our guidelines. We make code available at https://github.com/danini/affine-correspondences-for-camera-geometry.

1 Introduction

Estimating the geometric relationship between two images, such as homography or the epipolar geometry, is a fundamental step in computer vision approaches such as Structure-from-Motion [50], Multi-View Stereo [51], and SLAM [41].

Traditionally, geometric relations are estimated from point correspondences (PCs) between the two images [27, 42]. This ignores that correspondences are often rather established between image regions than between individual points to which the descriptors are finally often assigned. These regions, *i.e.* patches extracted around keypoints found by detectors such as DoG [31] or MSER [34], are oriented and have a specified size. Thus, they provide an affine transformation mapping feature regions to each other for each match [31, 44].

Many works [3–5, 8, 16, 24, 25, 29, 43, 45, 48] used this additional information provided by affine correspondences (ACs), *i.e.*, region-to-region matches, to design minimal solvers for camera geometry estimation. As each correspondence carries more information, such solvers require fewer matches than their traditional point-based counterparts. For example, only three affine correspondences are required to estimate the fundamental matrix [8] compared to seven point correspondences [27]. This increases the probability of drawing an all-inlier sample, thus decreasing the required number of RANSAC [17] iterations. Also, ACs are known to be more robust against view changes than point correspondences [38].

In terms of noise on the measurements, affine solvers are affected differently than their point-based counterparts. If the points are well-spread in the images, the amount of noise is small compared to the distances between the points. In this case, the influence of the noise on the solution computed by a minimal solver is relatively small. In contrast, the comparatively small regions around the keypoints that define the affine features are much more affected by the same level of noise. As such, we observe that affine correspondence-based solvers are significantly less accurate than point correspondence-based ones if used naively. Yet, when explicitly modelling the impact of noise, we observe that affine solvers can achieve a similar level of accuracy as classical approaches while offering faster run-times. Based on our observations, we provide a practical guide for making affine correspondences work well in camera geometry computation.

Contribution. (**1**) We demonstrate how to use affine solvers to obtain accurate results at faster RANSAC run-times than achieved by pure point-based solvers. (**2**) We present strategies for all parts of the camera geometry estimation pipeline designed to improve the performance of affine solvers. This includes the refinement of the affinities defined by the features, rejection of samples based on cheirality checks, uncertainty propagation to detect and reject models that are too uncertain, and the importance of local optimization. (**3**) Through detailed experiments, we evaluate the impact of each strategy on the overall performance of affine solvers, both in terms of accuracy and run-time, showing that affine solvers can achieve a similar or higher accuracy than point-based approaches at faster run-times. These experiments validate our guidelines. (**4**) We make various technical contributions, such as a novel method for refining affine correspondence based on image intensity; a new minimal solver for fundamental matrix estimation; a strategy for combining the SPRT [11] test with uncertainty propagation for rejecting models early; the adaptation of the sample cheirality test, which is often used for point-based homography estimation, to affine features, and a general scheme for deriving covariance matrices for minimal solvers.

2   Related Work

Our guide to best use affine correspondences for camera geometry estimation problems analyzes the individual stages of the pipeline leading from matches to transformation estimates. The following reviews prior work for each stage.

Affine features are described by a point correspondence and a 2×2 linear transformation. For obtaining them, one can apply one of the traditional affine-covariant feature detectors, thoroughly surveyed in [37], such as MSER, Hessian-Affine, or Harris-Affine detectors. An alternative way to acquiring affine features is via view-synthesizing, as done, *e.g.*, by Affine-SIFT [40], and MODS [39] or by learning-based approaches, *e.g.*, Hes-Aff-Net [38], which obtains affine regions by running CNN-based shape regression on Hessian keypoints.

Matching affine regions. Given the noise in the parameters of the regions around affine matches, a natural approach to good estimation is to use high precision least squares matching (LSM) for refining affine correspondences [1,18,33]. Similar to template matching via cross correlation, a small patch from one image is located within a larger region in a second image. While arbitrary geometric and radiometric models are possible, practical approaches mostly consider *affine transformations*. As a maximum likelihood estimator, LSM provides the *covariance matrix of the estimated parameters*, the Cramer-Rao bound, reaching standard deviations for parallaxes down to below $1/100$ pixel [26]. Intensity-based refinement has been used [16, 48] for pose estimation. Yet, no analysis on the accuracy of the derived uncertainty is known and a symmetric formulation of the problem is missing so far. In this paper, we close this gap by providing both.

Affine solvers use ACs for geometric model estimation. Bentolila and Francos [8] proposed a method for estimating the epipolar geometry between two views using three ACs by interpreting the problem via conic constraints. Perdoch et al. [43] proposed two techniques for approximating the pose based on two and three matches by converting each AC to three PCs and applying standard estimation techniques. Raposo et al. [48] proposed a solution for essential matrix estimation from two ACs. Baráth et al. [3, 5] showed that the relationship between ACs and epipolar geometry is linear and geometrically interpretable. Eichhardt et al. [16] proposed a method that uses two ACs for relative pose estimation based on general central-projective views. Similarly, [24,25] proposed minimal solutions for relative pose from a single affine correspondence when the camera is mounted to a moving vehicle. Homographies can also be estimated from two ACs as first shown by Köser [4,29]. Pritts et al. [45] used affine features for simultaneous estimation of affine image rectification and lens distortion.

Uncertainty analysis of image data provides several approaches useful for our task. Variances or covariance matrices are often used to model the *uncertainty of the input parameters*, *i.e.*, image intensities, coordinates of keypoints, and affine parameters. Assuming an ideal camera with a linear transfer function, the variance of the image noise increases linearly with the intensity [15,54]. In practice, the complexity of the internal camera processing requires an estimate of the variance function $\sigma_n^2(I)$ from the given images [19]. The accuracy of keypoint coordinates usually lies in the order of the rounding error, *i.e.*, $1/\sqrt{12} \approx 0.3$ pixels. We exploit here the uncertainty of the image intensities and keypoints for deriving realistic covariance matrices of the affine correspondences.

Propagation of input uncertainty to model parameters through the estimation depends on the model being estimated. The uncertainty of a homography

---

Algorithm 1 Robust model estimation pipeline with ACs

Require: I_1, I_2 – images
 1: $\mathcal{A} \leftarrow$ DetectACs(I_1, I_2) $\triangleright$ Sec. 4.1, *default*: SIFT desc. [32], DoG shape adapt. [31]
 2: $\widehat{\mathcal{A}} \leftarrow$ RefineACs$(\mathcal{A})$ $\triangleright$ Sec. 3.2, *default*: symmetric LSM refinement
 3: $\theta^*, q^* \leftarrow 0, 0$ $\triangleright$ Best model and its quality
 4: **while** ¬Terminate **do** $\triangleright$ Robust estimation, *default*: GC-RANSAC [6]
 5:      $S \leftarrow$ Sample$(\widehat{\mathcal{A}})$ $\triangleright$ *default*: PROSAC sampler [10]
 6:      **if** ¬ TestSample(S) **then** $\triangleright$ Sample degeneracy and cheirality tests, Sec. 3.4
 7:          continue
 8:      $\theta \leftarrow$ ModelEstimation(S) $\triangleright$ *default*: F – Sec. 3.3, E – [5], H – [4]
 9:      **if** ¬ TestModel(θ) **then** $\triangleright$ *default*: tests from USAC [46]
10:          continue
11:      **if** ¬ Preemption(θ) **then** $\triangleright$ Sec. 3.5, *default*: SPRT [11] + uncertainty test
12:          continue
13:      $q \leftarrow$ Validate$(\theta, \widehat{\mathcal{A}})$ $\triangleright$ Model quality calculation, *default*: MSAC score [55]
14:      **if** $q > q^*$ **then**
15:          $q^*, \theta^* \leftarrow q, \theta$
16:          $\theta' \leftarrow$ LocalOptimization$(\theta, \widehat{\mathcal{A}})$ $\triangleright$ *note*: only PCs are used from the ACs
17:          **if** ¬ TestModel(θ') **then** $\triangleright$ *default*: tests from USAC [46]
18:              continue
19:          $q' \leftarrow$ Validate$(\theta', \widehat{\mathcal{A}})$ $\triangleright$ Model quality calculation, *default*: MSAC score [55]
20:          **if** $q' > q^*$ **then**
21:              $q^*, \theta^* \leftarrow q', \theta'$

---

estimated from four or more points has been based on the SVD [13,47] and Lie groups [7]. The uncertainty of an estimated fundamental matrix has been also based on the SVD [53], but also on minimal representations [14]. Finally, the uncertainty for essential matrices has been derived using a minimal representation [21]. As far as we know, the propagation for affine solvers has not been presented before, and there was no general scheme for deriving covariance matrices for the solutions of minimal solvers. In this paper, we provide an efficient and general scheme and the uncertainty propagation for all minimal solvers used.

3   A Practical Guide to Using Affine Features

As argued in Sec. 1, and shown experimentally in Sec. 4.1, using ACs instead of point correspondences (PCs) leads to minimal solvers with smaller sample sizes but also to less accurate results. In the following, we analyze the individual stages of a classical matching pipeline and discuss how state-of-the-art results can be obtained with ACs. The pipeline is summarized in Alg. 1.

3.1   Definition of Affine Correspondences

Affine correspondences are defined in this paper as

$$\text{AC}: \quad \{\boldsymbol{y}_0, \boldsymbol{z}_0, {}^h\mathtt{A}\} \quad \text{with} \quad {}^h\mathtt{A} = \begin{bmatrix} \mathtt{A} & \boldsymbol{c} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \tag{1}$$
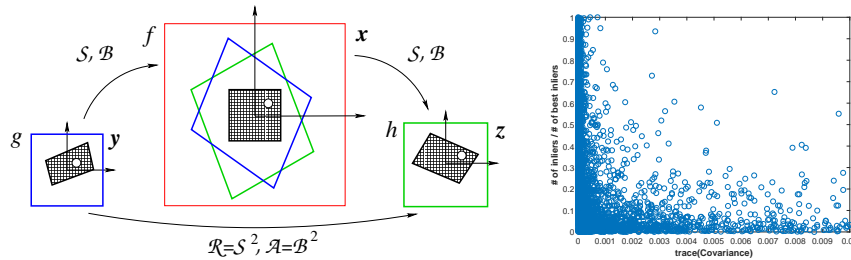
Fig. 1: **Left**: Relations between two image patches $g(\boldsymbol{y})$ (blue) and $h(\boldsymbol{h})$ (green) and the mean patch $f(\boldsymbol{x})$ (which is the black within the red region). The two patches g and h are related by geometric $\mathcal{B}$ and a radiometric $\mathcal{S}$ affinities. The correspondence is established by patch f, lying in the middle between g and h. We choose the maximum square (black), *i.e.* all the pixels in g and h which map into the black square of the reference image f. **Right**: The inlier ratio (vertical) of 100k homographies from real scenes as a function of the trace of their covariance matrices (horizontal). This shows that uncertain models (on right) generate small numbers of inliers. We use this to reject uncertain models.

where the keypoint coordinates in the left and the right image are $\boldsymbol{y}_0$, and $\boldsymbol{z}_0$, and the local affinity is $\mathtt{A} = \mathtt{A}_2\mathtt{A}_1^{-1}$, *e.g.* derived from the affine frames $\mathtt{A}_1$ and $\mathtt{A}_2$ representing the affine shape of the underlying image region. The matching and refinement of the affine region correspondences, presented in the next section, refers to the homogeneous matrix $^h\mathtt{A}$, specifically the translation vector $\boldsymbol{c}$, initially $\boldsymbol{0}$, and the affinity matrix $\mathtt{A}$ in (1).

3.2   Matching and Refining Affine Correspondences

For refining the ACs, we propose an intensity-based matching procedure which (i) is symmetric and provides (ii) a statistic for the coherence between the data and the model and (iii) a covariance matrix for the estimates of the parameters. Let the two image windows in the two images be $g(\boldsymbol{y})$ and $h(\boldsymbol{z})$. We assume, both windows are noisy observations of an unknown underlying signal $f(\boldsymbol{x})$, with individual geometric distortion, brightness, and contrast. We want to find the geometric distortion $\boldsymbol{z} = \mathcal{A}(\boldsymbol{y})$ and the radiometric distortion $h = \mathcal{R}(g) = pg+q$. Classical matching methods assume the geometric and radiometric distortion of one of the two windows is zero, *e.g.* $g(\boldsymbol{y}) = f(\boldsymbol{x})$, with $\boldsymbol{y} = \boldsymbol{x}$. We break this asymmetry by placing the unknown signal $f(\boldsymbol{x})$ in the middle between the observed signals g and h: $g(\boldsymbol{y}) \xrightarrow{\mathcal{S},\mathcal{B}} f(\boldsymbol{x}) \xrightarrow{\mathcal{S},\mathcal{B}} h(\boldsymbol{z})$ leading to $\mathcal{R} = \mathcal{S}^2$ and $\mathcal{A} = \mathcal{B}^2$. Assuming affinities for the geometric and radiometric distortions, the model is shown in Fig. 1(left). The geometric and the radiometric models are

$$\boldsymbol{x} = \mathtt{B}\boldsymbol{y} + b\,, \quad \boldsymbol{z} = \mathtt{B}\boldsymbol{x} + b \quad \text{and} \quad f = sg + t\,, \quad h = sf + t. \tag{2}$$

In the following, we collect the eight unknown parameters of the two affinities in a single vector $\boldsymbol{\theta} = [b_{11}, b_{21}, b_{12}, b_{22}, b_1, b_2, s, t]^\mathsf{T}$.

Now, we assume the intensities g and h to be noisy with variances $\sigma_n(g)$ and $\sigma_m(h)$ depending on g and h. Hence, in an ML-approach, we minimize the weighted sum $\Omega(\boldsymbol{\theta}, f) = \sum_j n_j^2(\boldsymbol{\theta}, f)/\sigma_{n_j}^2 + \sum_k m_k^2(\boldsymbol{\theta}, f)/\sigma_{m_k}^2$ w.r.t. the unknown parameters $\boldsymbol{\theta}$ and f, where the residuals are $n_j(\boldsymbol{\theta}, f) = g_j - s^{-1}\left(f\left(\mathtt{B}\boldsymbol{y}_j + \boldsymbol{b}\right) - t\right)$ and $m_k(\boldsymbol{\theta}, f) = h_k - \left(sf\left(\mathtt{B}^{-1}(\boldsymbol{z}_k - \boldsymbol{b})\right) + t\right)$. Since the number of intensities in the unknown signal f is quite large, we solve this problem by fixing one group of parameters of f and $\boldsymbol{\theta}$, and solving for the other. The estimated unknown function is the weighted mean of functions g and h transformed into the coordinate system $\boldsymbol{x}$ of f. The covariance matrix $\widehat{\Sigma}_{\widehat{\alpha}\widehat{\alpha}}$ of the sought affinity is finally derived by variance propagation from $\mathcal{A} = \mathcal{B}^2$. The standard deviations of the estimated affinity $\widehat{\mathtt{A}}$ and shift $\widehat{\boldsymbol{c}}$ are below 1% and 0.1 pixels, except for very small scales. Moreover, for the window size $M \times M$, the standard deviations decrease with on average with M^2 and M, respectively (see Supplementary Material).

3.3   Solvers using Affine Correspondences

In this paper we consider three important camera geometry problems: estimating planar homography, and two cases of estimating relative pose of two cameras: uncalibrated and calibrated cameras. We also include the semi-calibrated case, *i.e.* unknown focal length, in the supplementary material.

Homography from 1AC + 1PC: The problem of estimating a planar homography $\mathtt{H} \in \mathbb{R}^{3\times3}$ is well-studied with simple linear solutions from point and/or affine correspondences. The homography $\mathtt{H}$ has eight degrees of freedom. Since each PC gives two linear constraints on $\mathtt{H}$ and each AC gives six linear constraints on $\mathtt{H}$, the minimal number of correspondences necessary to estimate the unknown homography is either 4PC or 1AC+1PC. Both the well-known 4PC [27] and the 1AC+1PC [4, 29] solvers are solving a system of eight linear equations in nine unknowns and are therefore equivalent in terms of efficiency.

Fundamental matrix from 2AC + 1PC: The problem of estimating the relative pose of two uncalibrated cameras, *i.e.*, estimating the fundamental matrix $\mathtt{F} \in \mathbb{R}^{3\times3}$, has a well-known 7PC solver [27]. The fundamental matrix $\mathtt{F}$ has seven degrees of freedom, since it is a 3×3 singular matrix, *i.e.*, $\det(\mathtt{F}) = 0$. The well-known epipolar constraint gives one linear constraint on $\mathtt{F}$ for each PC. We propose a solver for estimating the unknown $\mathtt{F}$ using the linear constraints proposed in [3]. Here, we briefly describe this new 2AC+1PC solver.

Each AC gives three linear constraints on the epipolar geometry [3]. Therefore, the minimal number of correspondences necessary to estimate the unknown $\mathtt{F}$ is 2AC+1PC. The solver first uses seven linear constraints, *i.e.*, six from two ACs and one from a PC, rewritten in a matrix form as $\mathtt{M}\mathbf{f} = \mathbf{0}$, where $\mathbf{f} = \text{vec}(\mathtt{F})$, to find a 2-dimensional null-space of the matrix $\mathtt{M}$. The unknown fundamental matrix is parameterized as $\mathtt{F} = x\mathtt{F}_1 + \mathtt{F}_2$, where $\mathtt{F}_1$ and $\mathtt{F}_2$ are matrices created from the 2-dimensional null-space of $\mathtt{M}$ and x is a new unknown. This parameterization is substituted into the constraint $\det(\mathtt{F}) = 0$, resulting in a polynomial of degree three in one unknown. The final 2AC+1PC solver is performing the same operations as the 7PC solver, *i.e.*, the computation of the null-space of a 7×9 matrix and finding the roots of a univariate polynomial of degree three.

Essential matrix from 2ACs: The problem of estimating the relative pose of two calibrated cameras, *i.e.*, estimating the unknown essential matrix $\mathsf{E} \in \mathbb{R}^{3\times3}$, has five degrees of freedom (three for rotation and two for translation) and there exists the well-known 5PC solver [42]. This problem has recently been solved from two ACs [5,16]. Each AC gives three linear constraints on E [5]. Thus, two ACs provide more constraints than degrees of freedom. One approach to solve for E from two ACs is to use just five out of six constraints, which results in the same operations as the well-known 5PC solver [42] does. Another one is to solve an over-constrained system as suggested in [5]. In the experiments, we used the solver of [5] since it has lower computational complexity and similar accuracy.

3.4   Sample Rejection via Cheirality Checks

It is well-known for homography fitting that some minimal samples can be rejected without estimating the implied homography as they would lead to impossible configurations. Such a configuration occurs when the plane flips between the two views, *i.e.*, the second camera sees it from the back. This cheirality constraint is implemented in the most popular robust approaches, *e.g.*, USAC [46] and OpenCV's RANSAC. We thus adapt this constraint via a simple strategy by converting each AC to three PCs. Given affine correspondence $\{\boldsymbol{y}_0, \boldsymbol{z}_0, \mathsf{A}\}$, where $\boldsymbol{y}_0 = [y_{01}, y_{02}]^\mathsf{T}$ and $\boldsymbol{z}_0 = [z_{01}, z_{02}]^\mathsf{T}$ are the keypoint coordinates in the left, respectively the right images, the generated point correspondences are $(\boldsymbol{y}_0 + [1,0]^\mathsf{T}, \boldsymbol{z}_0 + \mathsf{A}[1,0]^\mathsf{T})$ and $(\boldsymbol{y}_0 + [0,1]^\mathsf{T}, \boldsymbol{z}_0 + \mathsf{A}[0,1]^\mathsf{T})$. When estimating the homography using the 1AC+1PC solver, the affine matrix is converted to these point correspondences and the cheirality check is applied to the four PCs.

Note that any direct conversion of ACs to (non-colinear) PCs is theoretically incorrect since the AC is a local approximation of the underlying homography [4]. However, it is a sufficiently good approximation for the cheirality check.

3.5   Uncertainty-based Model Rejection

Before evaluating the consensus of a model, it is reasonable to check its quality, especially to eliminate configurations close to a singularity, see [22]. We can use the covariance matrix $\Sigma_{\theta\theta}$ of each solution to decide on its further suitability. To do so, we propose a new general way of deriving the $\Sigma_{\theta\theta}$ for minimal problems.

All problems we address here are based on a set of constraints $\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta}) = \boldsymbol{0}$ on some observations $\boldsymbol{y}$, and parameters $\boldsymbol{\theta}$, *e.g.* the F estimation constraint $\boldsymbol{g}_i(\boldsymbol{y}_i, \boldsymbol{\theta})$ is of the form $\mathbf{x}_i^\mathsf{T}\mathsf{F}\mathbf{y}_i = 0$, hence $(\boldsymbol{y}, \boldsymbol{\theta}) = ([\mathbf{x}; \mathbf{y}], \text{vec}(\mathsf{F}))$. We want to use classical variance propagation for implicit functions [21], Sec. 2.7.5. If Σ_{yy} is given, the determination of $\Sigma_{\theta\theta}$ is based on linearizing $\boldsymbol{g}$ at a point $(\boldsymbol{y}, \boldsymbol{\theta})$ and using the Jacobians $\mathsf{A} = \partial\boldsymbol{g}/\partial\boldsymbol{y}$ and $\mathsf{B} = \partial\boldsymbol{g}/\partial\boldsymbol{\theta}$ leading to covariance matrix $\Sigma_{\theta\theta} = \mathsf{B}^{-1}\mathsf{A}\Sigma_{yy}\mathsf{A}^\mathsf{T}\mathsf{B}^{-\mathsf{T}}$, provided B can be inverted. Using constraints $\boldsymbol{g}$ we derive Jacobians A, B algebraically. Further, given the kth solution of a minimal problem (a system of equations defining a minimal problem has, in general, more than one solution), *i.e.* a pair $(\boldsymbol{y}, \boldsymbol{\theta}_k)$, we can compute $\Sigma_{\theta_k\theta_k}$ without needing to know how the problem was solved and how this specific solution has been selected.

However, the number of constraints in $\boldsymbol{g}$ in most of the minimal problems is smaller than the number of parameters $\boldsymbol{\theta}$, *e.g.* 7 constraints vs. 9 elements of F. Then the matrix B cannot be inverted. We propose to append the minimum number of constraints $\boldsymbol{h}(\boldsymbol{\theta}) = \boldsymbol{0}$ (between the parameters only, *e.g.* $\det(\mathtt{F})$ and $\|\mathtt{F}\| = 1$) such that the number of all constraints $(\boldsymbol{g}; \boldsymbol{h})$ is identical to the number of parameters. This leads to a regular matrix B, except for critical configurations.

Such algebraic derivations should be checked to ensure the equivalence of the algebraic and numerical solution, best by Monte Carlo simulations. However, the empirically obtained covariance matrix $\widehat{\Sigma}_{\theta\theta}$ is regular and cannot be directly compared to the algebraically derived $\Sigma_{\theta\theta}$ if it is singular (*e.g.* for $\boldsymbol{\theta} = \text{vec}(\mathtt{F})$). We propose to project both matrices on the tangent space of the manifold of $\boldsymbol{\theta}$, leading to regular covariance matrices as follows: let $\mathtt{J}(\Sigma)$ be an orthonormal base of the column space of some covariance matrix Σ, then $\Sigma_r = \mathtt{J}^\mathsf{T}\Sigma\mathtt{J}$ is regular; *e.g.* using $\mathtt{J} = \mathtt{null}(\mathtt{null}(\Sigma)^\mathsf{T})$ where $\mathtt{null}(\Sigma)$ is the nullspace of Σ. Hence, with $\mathtt{J} = \mathtt{J}(\Sigma_{\theta\theta})$ the two covariance matrices $\Sigma_{\theta\theta,r} = \mathtt{J}^\mathsf{T}\Sigma_{\theta\theta}\mathtt{J}$ and $\widehat{\Sigma}_{\theta\theta,r} = \mathtt{J}^\mathsf{T}\widehat{\Sigma}_{\theta\theta}\mathtt{J}$ are regular. They can be compared and an identity test can be performed checking the hypothesis $\mathbb{E}(\widehat{\Sigma}_{\theta\theta,r}) = \Sigma_{\theta\theta,r}$, see [21], p. 71. The used constraints and detailed discussion for all listed minimal problems are in the supplementary material.

The covariance matrix $\Sigma_{\theta\theta}$ can be used in following way. Since we do not have a reference configuration, we eliminate models where the condition number $c = \text{cond}(\Sigma_{\theta\theta})$ is too large, since configurations close to singularity show large condition numbers in practice. For reasons of speed, it is useful to compute an approximation for the condition number, *e.g.* $c_s = \text{tr}(\Sigma_{\theta\theta})\text{tr}(\Sigma_{\theta\theta}^{-1})$, if the inverse covariance matrix can be obtained efficiently, which is the case in our context since $\Sigma_{\theta\theta}^{-1} = \mathtt{B}(\mathtt{A}\Sigma_{yy}\mathtt{A}^\mathsf{T})^{-1}\mathtt{B}^\mathsf{T}$.[6] A weaker measure is $\text{tr}(\Sigma_{\theta\theta})$, which is more efficient to calculate than the previous ones. It essentially measures the average variance of the parameters $\boldsymbol{\theta}$. Thus, it can identify configurations where parameters are very uncertain. We use this measure in the following for deriving a prior for preemptive model verification by the Sequential Probability Ratio Test [11].

We experimentally found, for each problem, the parameters of exponential (for points solvers) and log-normal (for affine solvers) distributions for the trace. These parameters are used to measure the likelihood of the model being too uncertain to lead to a large number of inliers. In our experiments, for the sake of simplicity, we model the trace values by normal distributions for all solvers. Thus, we used the a-priori determined parameters to initialize the mean and standard deviation from all tested image pairs. Finally, we get a probability for each model being acceptable or not. Note that the selection of the provably correct probabilistic kernel for a particular problem and scene is a direction for future research. However, it is not a crucial issue due to being used only for rejecting too uncertain models early to avoid unnecessary calculations.

As a final step, we feed the determined probability to the Sequential Probability Ratio Test (SPRT) [11,35] as a prior knowledge about the model to be

---

[6] For $\Sigma = \text{Diag}([\text{a,b}])$, with $a > b$, the condition number is $c = a/b$, while the approximation is $c_s = (a + b)^2/(ab)$, which for $a \gg b$ converges to the condition number.

verified. This is done by initializing the model probability, which is sequentially updated in SPRT, to the a priori estimated one.

3.6   Local Optimization

Minimal solvers do not take noise in their input measurements into account during the estimation process. However, noise affects the estimated models. As such, not every all-inlier sample leads to the best possible transformation [12]. As shown in [20, 49], starting from the algebraic solution and performing only a single iteration of ML estimation is often sufficient to obtain a significantly better estimate. They show that this strategy approaches the optimal result with an error below 10%-40% of the parameters' standard deviations while only increasing the computation time by a factor of ∼2. A strongly recommended approach is thus to use local optimization [6,12,30] inside RANSAC: every time a new best model is found, ML-based refinement on its inliers is used to account for noise in the input parameters. While this adds a computational overhead, it can be shown that this overhead is small and is compensated by the observation that local optimization (LO) typically helps RANSAC to terminate early. Moreover LO is usually applied rarely [12]. As we show in Sec. 4.5, local optimization is crucial to obtain accurate geometry estimates when using ACs.

4   Experiments

In this section, different algorithmic choices are tested on homography, fundamental and essential matrix fitting problems to provide a pipeline which leads to results superior to point-based methods.

Experimental setup. Tests for epipolar geometry estimation were performed on the benchmark of [9]. The used datasets are the TUM dataset [52] consisting of videos, of resolution 640×480, of indoor scenes. The KITTI dataset [23] consists of consecutive frames of a camera mounted to a vehicle. The images are of resolution 1226×370. Both KITTI and TUM have image pairs with short baselines. The Tanks and Temples dataset [28] provides images of real-world objects for image-based reconstruction and, thus, contains mostly wide-baseline pairs. The images are of sizes between 1080×1920 and 1080×2048. The benchmark provides 1 000 image pairs for each dataset with ground truth epipolar geometry. Homography estimation was tested on the scenes of the HPatches dataset [2]. RANSAC's inlier-outlier threshold is set to 1.0 px (F), 1.0 px (E) and 5 px (H).

When evaluating F and E matrices, we calculate the normalized symmetric geometry errors (NSGD). The symmetric geometry error (SGD) was proposed in [58]. It generates virtual correspondences using the ground-truth F and computes the epipolar distance to the estimated F. It then reverts their roles to compute symmetric distance. The SGD error (in pixels) causes comparability issues for images of different resolutions. Therefore, it is normalized into the range of $[0, 1]$ by regularizing by factor $f = \frac{1}{\sqrt{h^2 + w^2}}$, where h and w are the height and width of the image, respectively.

The error of the estimated homographies is measured by, first, projecting the first image to the second one and back to get the commonly visible area. Each pixel in the visible area is projected by the ground truth and estimated homographies and the error is calculated as the L_2 distance of the projected points. Finally, the error is averaged over all pixels of the visible area.

4.1   Matchers and descriptors

To estimate ACs in real images, we applied the VLFeat library [56] since it is available for multiple programming languages and, thus, we considered it a practical choice. VLFeat provides several options either for the feature descriptor or the affine shape adaptation technique. To select the best-performing combination, first, we detected ACs using all of the possible combinations. Note that we excluded the multi-scale versions of Harris-Laplace and Hessian-Laplace [36] affine shape adaptations since they were computationally expensive. Correspondences are filtered by the standard second nearest neighbor ratio test [32]. Next, we estimated fundamental matrices using affine and point-based solvers and *vanilla* RANSAC [17]. Fig. 2 reports the cumulative distribution function (CDF) of the NSGD errors calculated from the estimated fundamental matrices.

Curves showing affine solvers have circles as markers. We applied PC-based methods (crosses) considering only the locations of the correspondences and ignoring the affinities. The line style denotes the feature descriptor: straight line – SIFT [32], dotted – LIOP [57]. Affine shape adaption techniques (DoG, Hessian [37], Harris- and Hessian-Leplace [36]) are shown by color. Applying VLFeat with any affine shape adaptation increases the extraction time by ≈10%.

The first and most dominant trend visible from the plots is that methods exploiting ACs are significantly less accurate then point-based approaches when the naive approach is used: vanilla RANSAC. The SIFT descriptor [32] and DoG affine shape adaptation lead to the most accurate results. Consequently, we use this combination in the experiments. In the next sections we will show ways of making the affine solvers similarly or more accurate than point-based methods.

4.2   Match Refinement

We demonstrate how the proposed refinement affects the accuracy of the affine matches by analysing the fulfillment of the constraints in the case of calibrated cameras. For each AC, the three constraints consist of the epipolar constraint $c_p = \boldsymbol{c}_p^\mathsf{T} \boldsymbol{e} = 0$ for the image coordinates, and the two constraints $\boldsymbol{c}_a = \mathsf{C}_a^\mathsf{T} \boldsymbol{e} = \boldsymbol{0}$ for the affinity A. Assuming the pose, *i.e.* the essential matrix, is known we determine a test statistics for the residuals c_p and $\boldsymbol{c}_a$ as $d_p = ||c_p||_{\sigma_{c_p}^2}$ and $d_a = ||\boldsymbol{c}_a||_{\Sigma_{c_a c_a}}$. For the ACs of five image pairs, we used the LOWE keypoint coordinates and the scale and direction differences for deriving approximate affinities and refined them using the proposed LSM refinement technique (see Section 3.2).

Fig. 3, left shows the CDF of the improvement caused by the proposed technique in the point coordinates (r_p) and in the affine parameters (r_a); a bigger
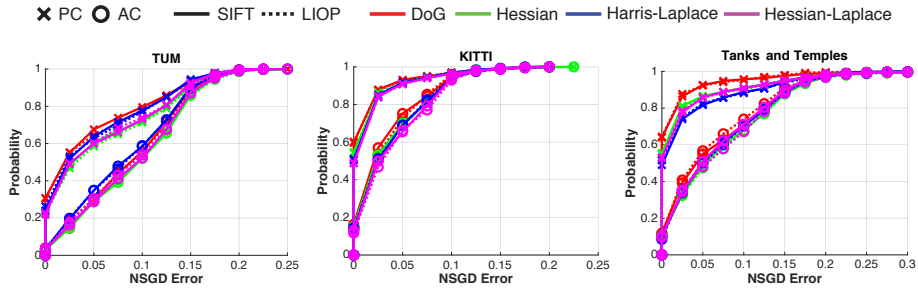
Fig. 2: Fundamental matrix estimation on datasets TUM, KITTI and Tanks and Temples (from benchmark [9]; 1000 image pairs each) using ACs detected by different descriptors and detectors. The CDFs of NSGD errors are shown. Vanilla RANSAC was applied followed by a LS fitting on all inliers.

value is better. The method improves both the affine parameters and point co-ordinates significantly. In Fig. 3, right the inconsistency with the epipolar con-straints are shown; smaller values are better. The refined ACs are better, in terms of fulfilling the epipolar constraints, than the input ACs.
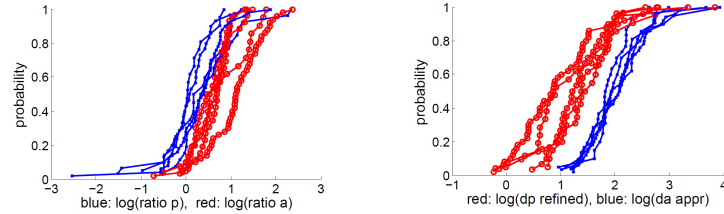
Fig. 3: **Left:** Improvement of the points (r_p) and affine parameters (r_a) after the proposed refinement, larger values are better. **Right:** Inconsistency with the epipolar constraints before (blue) and after (red) the refinement, smaller values are better. The CDFs are calculated from five images.

4.3   Sample Rejection via Cheirality Checks

The widely used technique for rejecting minimal samples early (*i.e.* without estimating the model parameters) when fitting homographies is the ordering check of the point correspondences as described earlier. Its effect when adapting it to affine correspondences is shown in Fig. 4.

In the left plot of Fig. 4, the cumulative distribution functions of the pro-cessing times (in seconds) are shown. It can be seen that this test has a huge effect on point-based homography estimation as it speeds up the procedure signif-icantly. The adapted criterion speeds up affine-based estimation as well, however,
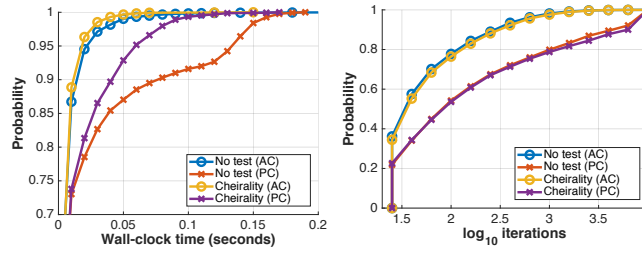
Fig. 4: Effect of cheirality test on point and affine-based homography estimators.

not that dramatically. Note, that affine-based estimation is already an order of magnitude faster than point-based methods and for AC-based homography estimation the cumulative distribution curve of the processing time is already very steep, Fig. 4 (left). This means that affine-based estimators do not perform many iterations and skipping model verification for even the half of the cases would not affect the time curve significantly. The avg. processing time of affine-based estimation is dropped from 8.6 to 7.7 ms. The right plot shows the $\log_{10}$ iteration numbers of the methods. The test does not affect the iteration number significantly. It sometimes leads to more iterations due to not checking samples of impossible configurations, however, this is expected and does not negatively affect the time.

4.4   Uncertainty-Based Preemptive Verification

(a) Homographies.   (b) Fundamental matrices.   (c) Essential matrices.
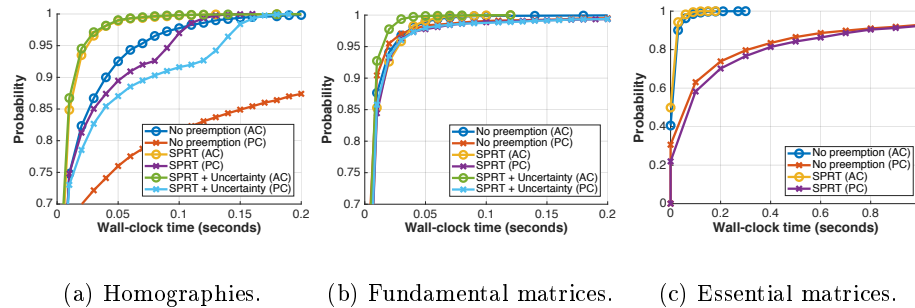
Fig. 5: Evaluating pre-emptive model verification strategies for affine (AC) and point-based (PC) robust estimation. The CDFs of the processing times (in seconds) are shown. Being fast is interpreted as a curve close to the top-left corner.

As it is described earlier, we combined the SPRT test [11, 35] (parameters are set similarly as in the USAC [46] implementation) with model uncertainty

calculation to avoid the expensive validation of models which are likely to be worse than the current best one. Fig. 5 reports the CDFs of the processing time for (a) homography, (b) fundamental and (c) essential matrix fitting. Note that we excluded uncertainty-based verification for essential matrices since the solvers became too complex and, thus, the uncertainty calculation was slow for being applied to every estimated model.

It can clearly be seen that the proposed combination of the SPRT and the uncertainty check leads to the fastest robust estimation both for H and F fitting. For E estimation from ACs, using the SPRT test is also important, leading to faster termination. Most importantly, using affine correspondences, compared to point-based solvers, leads to a significant speed-up for all problems.

4.5   The Importance of Local Optimization

While the speed-up caused by using ACs has clearly been demonstrated, the other most important aspect is to get accurate results. As it is shown in Fig. 2, including AC-based solvers in *vanilla* RANSAC leads to significantly less accurate results than using PCs. A way of making the estimation by ACs accurate is to use a locally optimized RANSAC, where the initial model is estimated by an AC-based minimal solver and the local optimization performs the model polishing solely on the inlier PCs. We tested state-of-the-art local optimization techniques, *i.e.*, LO-RANSAC [12], LO'-RANSAC [30], GC-RANSAC [6]. The results are reported in Fig. 6. It can be seen that affine-based estimation with GC-RANSAC is always among the top-performing methods. In (a), it is marginally more accurate than considering only point correspondences. In (b), using point correspondences is slighly more accurate. Compared to the results of *vanilla* RANSAC, the results of affine-based robust estimation improved notably.

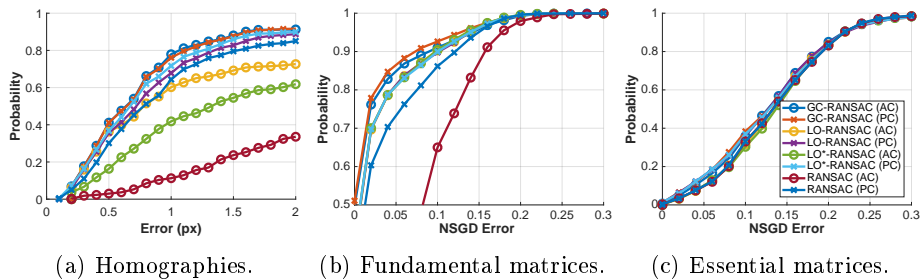(a) Homographies.   (b) Fundamental matrices.   (c) Essential matrices.

Fig. 6: Evaluating local optimization techniques for affine (AC) and point-based (PC) robust model estimation. The CDFs of the geometric errors are shown. Being accurate is interpreted as a curve close to the top-left corner.

5   Discussion

In summary of the investigated approaches, the best practices to accurately and efficiently use affine solvers are the following. Affinity-based model estimation has an accuracy similar or better to point-based solvers if (i) the detector and affine shape refining method is carefully selected; and (ii), most importantly, if a locally optimized RANSAC is applied to polish every new so-far-the-best model using *only* the point locations from the inlier correspondences. Consequently, affine features are used for estimating models from minimal samples, while their point counterparts are used to obtain accurate results. Efficiency is achieved by (iii) adapting strategies well-established for point correspondences, *e.g.*, cheirality check for homography estimation. (iv) Also, uncertainty-based model rejection and other preemptive verification techniques have a significant impact when speeding up the robust estimation procedure.

6   Conclusions

In this paper, we have considered the problem of using affine correspondences (ACs) for camera geometry estimation, *i.e.*, homography and epipolar geometry computation. Compared to classical approaches based on point correspondences (PCs), minimal solvers based on ACs offer the advantage of smaller sample sizes and, thus, the possibility to significantly accelerate RANSAC-based randomized robust estimation. However, noise has a larger negative impact on affine solvers as their input measurements typically originate from a smaller image region compared to point solvers. As we have shown, this significantly decreases the accuracy of the affine solvers. In this work, we have thus collected a set of "best practices", including novel contributions such as refining the local feature geometry and uncertainty-based model rejection techniques, for using ACs in practice. Through extensive experiments, we have shown that following our guidelines enables affine solvers to be used effectively, resulting in similar accuracy but faster run-times compared to point-based solvers. We believe that our guide will be valuable for both practitioners aiming to improve the performance of their pipelines as well as researchers working on ACs as it covers a topic previously unexplored in the literature.

References

1. Ackermann, F.: Digital image correlation: Performance and potential application in photogrammetry. Photogrammetric Record **11**(64), 429–439 (1984)
2. Balntas, V., Lenc, K., Vedaldi, A., Mikolajczyk, K.: HPatches: A benchmark and evaluation of handcrafted and learned local descriptors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5173–5182 (2017)
3. Baráth, D., Tóth, T., Hajder, L.: A minimal solution for two-view focal-length estimation using two affine correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)
4. Barath, D., Hajder, L.: A theory of point-wise homography estimation. Pattern Recognition Letters **94**, 7–14 (2017)
5. Barath, D., Hajder, L.: Efficient recovery of essential matrix from two affine correspondences. IEEE Transactions on Image Processing **27**(11), 5328–5337 (2018)
6. Barath, D., Matas, J.: Graph-cut RANSAC. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6733–6741 (2018)
7. Begelfor, E., Werman, M.: How to put probabilities on homographies. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1666–1670 (2005), http://dx.doi.org/10.1109/TPAMI.2005.200
8. Bentolila, J., Francos, J.M.: Conic epipolar constraints from affine correspondences. Computer Vision and Image Understanding **122**, 105–114 (2014)
9. Bian, J.W., Wu, Y.H., Zhao, J., Liu, Y., Zhang, L., Cheng, M.M., Reid, I.: An evaluation of feature matchers forfundamental matrix estimation. arXiv preprint arXiv:1908.09474 (2019), https://jwbian.net/fm-bench
10. Chum, O., Matas, J.: Matching with PROSAC-progressive sample consensus. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. vol. 1, pp. 220–226. IEEE (2005)
11. Chum, O., Matas, J.: Optimal randomized RANSAC. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(8), 1472–1482 (2008)
12. Chum, O., Matas, J., Kittler, J.: Locally optimized RANSAC. In: Joint Pattern Recognition Symposium. pp. 236–243. Springer (2003)
13. Criminisi, A.: Accurate Visual Metrology from Single and Multiple Uncalibrated Images. Springer (2001)
14. Csurka, G., Zeller, C., Zhang, Z., Faugeras, O.D.: Characterizing the uncertainty of the fundamental matrix. Computer vision and image understanding **68**(1), 18–36 (1997)
15. Dainty, J.C., Shaw, R.: Image Science. Academic Press (1974)
16. Eichhardt, I., Chetverikov, D.: Affine correspondences between central cameras for rapid relative pose estimation. In: Proceedings of the European Conference on Computer Vision. pp. 482–497 (2018)
17. Fischler, M., Bolles, R.: Random Sampling Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. Commun. ACM **24**, 381–395 (1981)
18. Förstner, W.: On the geometric precision of digital correlation. In: Hakkarainen, J., Kilpelä, E., Savolainen, A. (eds.) Intl. Archives of Photogrammetry. vol. XXIV-3, pp. 176–189. ISPRS Symposium, Comm. III, Helsinki (Jun 1982), http://www.ipb.uni-bonn.de/pdfs/Forstner1982Geometric.pdf

19. Förstner, W.: Image preprocessing for feature extraction in digital intensity, color and range images. In: Geomatic Methods for the Analysis of Data in Earth Sciences, Lecture Notes in Earth Sciences, vol. 95/2000, pp. 165–189. Springer (2000), http://www.ipb.uni-bonn.de/pdfs/Forstner2000Image.pdf
20. Förstner, W., Khoshelham, K.: Efficient and accurate registration of point clouds with plane to plane correspondences. In: 3rd International Workshop on Recovering 6D Object Pose (2017), http://www.ipb.uni-bonn.de/pdfs/foerstner17efficient.pdf
21. Förstner, W., Wrobel, B.P.: Photogrammetric Computer Vision. Springer (2016), http://www.ipb.uni-bonn.de/book-pcv/
22. Frahm, J.M., Pollefeys, M.: RANSAC for (quasi-)degenerate data (QDEGSAC). vol. 1, pp. 453–460 (07 2006)
23. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? The KITTI vision benchmark suite. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition. pp. 3354–3361. IEEE (2012)
24. Guan, B., Zhao, J., Li, Z., Sun, F., Fraundorfer, F.: Minimal solutions for relative pose with a single affine correspondence (2020)
25. Hajder, L., Baráth, D.: Relative planar motion for vehicle-mounted cameras from a single affine correspondence. In: IEEE International Conference on Robotics and Automation (2020)
26. Haralick, R., Shapiro, L.G.: Computer and Robot Vision, vol. II. Addison-Wesley, Reading, MA (1992), http://www.ipb.uni-bonn.de/pdfs/Forstner1993Image.pdf
27. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
28. Knapitsch, A., Park, J., Zhou, Q.Y., Koltun, V.: Tanks and Temples: Benchmarking large-scale scene reconstruction. ACM Transactions on Graphics **36**(4), 78 (2017)
29. Köser, K.: Geometric Estimation with Local Affine Frames and Free-form Surfaces. Shaker (2009)
30. Lebeda, K., Matas, J., Chum, O.: Fixing the locally optimized RANSAC. In: British machine vision conference. pp. 1–11. Citeseer (2012)
31. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal on Computer Vision **60**(2) (2004)
32. Lowe, D.G.: Object recognition from local scale-invariant features. In: International Conference on Computer vision. IEEE (1999)
33. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: IJCAI '81. pp. 674–679 (1981)
34. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing **22**(10), 761 – 767 (2004), http://www.sciencedirect.com/science/article/pii/S0262885604000435
35. Matas, J., Chum, O.: Randomized RANSAC with sequential probability ratio test. In: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1. vol. 2, pp. 1727–1732. IEEE (2005)
36. Mikolajczyk, K., Schmid, C.: Indexing based on scale invariant interest points. In: Proceedings Eighth IEEE International Conference on Computer Vision. vol. 1, pp. 525–531. IEEE (2001)
37. Mikolajczyk, K., Schmid, C.: An affine invariant interest point detector. In: European conference on computer vision. pp. 128–142. Springer (2002)
38. Mishkin, D., Radenovic, F., Matas, J.: Repeatability is not enough: Learning affine regions via discriminability. In: Proceedings of the European Conference on Computer Vision. pp. 284–300 (2018)

39. Mishkin, D., Matas, J., Perdoch, M.: Mods: Fast and robust method for two-view matching. Computer Vision and Image Understanding **141**, 81–93 (2015)
40. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. SIAM journal on imaging sciences **2**(2), 438–469 (2009)
41. Mur-Artal, R., Tardós, J.D.: ORB-SLAM2: an Open-Source SLAM System for Monocular, Stereo and RGB-D Cameras. TRO **33**(5), 1255–1262 (2017)
42. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE TPAMI pp. 756–770 (2004)
43. Perdoch, M., Matas, J., Chum, O.: Epipolar geometry from two correspondences. In: International Conference on Computer Vision (2006)
44. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object Retrieval with Large Vocabularies and Fast Spatial Matching. In: Conference on Computer Vision and Pattern Recognition (2007)
45. Pritts, J.B., Kukelova, Z., Larsson, V., Lochman, Y., Chum, O.: Minimal solvers for rectifying from radially-distorted conjugate translations. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
46. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.M.: USAC: a universal framework for random sample consensus. IEEE transactions on pattern analysis and machine intelligence **35**(8), 2022–2038 (2012)
47. Raguram, R., Frahm, J.M., Pollefeys, M.: Exploiting uncertainty in random sample consensus. In: International Conference on Computer Vision. pp. 2074–2081. IEEE (2009)
48. Raposo, C., Barreto, J.P.: Theory and practice of structure-from-motion using affine correspondences. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 5470–5478 (2016)
49. Schneider, J., Stachniss, C., Förstner, W.: On the quality and efficiency of approximate solutions to bundle adjustment with epipolar and trifocal constraints. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences. vol. IV-2/W3, pp. 81–88 (2017). https://doi.org/10.5194/isprs-annals-IV-2-W3-81-2017, https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/IV-2-W3/81/2017/isprs-annals-IV-2-W3-81-2017.pdf
50. Schönberger, J.L., Frahm, J.M.: Structure-From-Motion Revisited. In: CVPR (June 2016)
51. Schönberger, J.L., Zheng, E., Pollefeys, M., Frahm, J.M.: Pixelwise View Selection for Unstructured Multi-View Stereo. In: European Conference on Computer Vision (ECCV) (2016)
52. Sturm, J., Engelhard, N., Endres, F., Burgard, W., Cremers, D.: A benchmark for the evaluation of RGB-D SLAM systems. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp. 573–580. IEEE (2012)
53. Sur, F., Noury, N., Berger, M.O.: Computing the uncertainty of the 8 point algorithm for fundamental matrix estimation (2008)
54. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer (2010)
55. Torr, P.H.S.: Bayesian model estimation and selection for epipolar geometry and generic manifold fitting. IJCV (2002)
56. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. In: Proceedings of the 18th ACM international conference on Multimedia. pp. 1469–1472 (2010)
57. Wang, Z., Fan, B., Wu, F.: Local intensity order pattern for feature description. In: 2011 International Conference on Computer Vision. pp. 603–610. IEEE (2011)
58. Zhang, Z.: Determining the epipolar geometry and its uncertainty: A review. International journal of computer vision **27**(2), 161–195 (1998)