Making Affine Correspondences Work in Camera Geometry Computation - Supplementary Material

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

This supplementary material provides the following information: Sec. 1 provides technical details for the matching approach described in Sec. 3.2 in the paper. Sec. 2 describes a solver that estimates the essential matrix and a common focal length from affine correspondences (ACs) (*c.f.* Sec. 3.3 in the paper). Sec. 3 provides details on the constraints used to derive uncertainty estimates for the solvers used in the paper (*c.f.* Sec. 3.5 in the paper). Sec. 4 discusses fitting a distribution other than the normal distribution to the trace of the covariance matrix (*c.f.* Sec. 3.5 in the paper). Sec. 5 finally provides additional results on the impact of local optimization (*c.f.* Sec. 4.5 in the paper).

1 Symmetric Least Squares Matching

Here we describe the symmetric version of least squares matching in more detail and give some experimental results.

1.1 Model

Let the two image windows $g(\boldsymbol{y})$ and $h(\boldsymbol{z})$ in the two images be given (*c.f.* Fig. 1). The coordinates refer to the centre of the square windows. We assume that both windows are noisy observations of an unknown true underlying signal $f(\boldsymbol{x})$, with individual geometric distortion, brightness, and contrast. We want to determine 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.* assuming $g(\boldsymbol{y}) = f(\boldsymbol{x})$, with $\boldsymbol{y} = \boldsymbol{x}$. We break this asymmetry by placing the unknown signal $f(\boldsymbol{x})$ in the middle between the observed signals between g and h:

$$g(\boldsymbol{y}) \overset{\mathcal{B},\mathcal{S}}{\longrightarrow} f(\boldsymbol{x}) \overset{\mathcal{B},\mathcal{S}}{\longrightarrow} h(\boldsymbol{z}) \quad \text{such that} \quad \mathcal{A} = \mathcal{B}^2 \,, \mathcal{R} = \mathcal{S}^2 \,. \tag{1}$$

Assuming affinities for the geometric and the radiometric distortion, we have the following generative model (see Fig. 1): The geometric and the radiometric

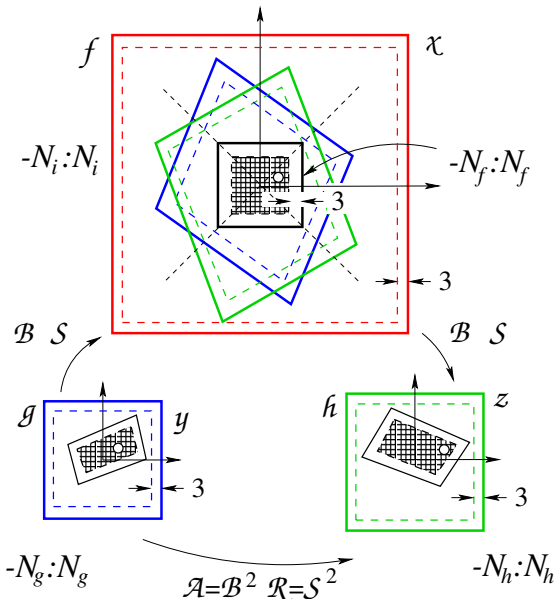


Fig. 1: Relations between two given square image patches $g(\boldsymbol{y})$ (blue) and $h(\boldsymbol{h})$ (green) and the mean patch $f(\boldsymbol{x})$ (which is the black within the red region). The two image patches g and h are related by geometric and a radiometric affinities B and S, respectively. The correspondence is established by the patch f. Geometrically and radiometrically it lies in the middle between g and h. Only a region in the overlap of the two patches g and h mapped to f can be used. We choose the maximum square (black). The observations are all pixels in g and h which map into the black square of the reference image f. We assume the reference image f is a restored version of the weighted mean of the two projected images g and h. The patches g and h may have different sizes. The size of the unknown signal (black, gray) depends on the sizes of g and h, the approximate affine transformation A and a border to allow bicubic interpolation, and is adapted in each iteration. The large image in the $\boldsymbol{x}$-frame is used for generating artificial images. The dashed lines indicate the borders required for allowing bicubic interpolation

models for the two images are

$$\boldsymbol{y} \mapsto \boldsymbol{x}: \quad \boldsymbol{x} = \mathtt{B}\boldsymbol{y} + \boldsymbol{b} \quad \text{and} \quad \boldsymbol{x} \mapsto \boldsymbol{z}: \quad \boldsymbol{z} = \mathtt{B}\boldsymbol{x} + \boldsymbol{b} \tag{2}$$

$$g \mapsto f: \quad f = sg + t \quad \text{and} \quad f \mapsto h: \quad h = sf + t\,. \tag{3}$$

In the following we collect the eight unknown parameters of the affinities $\mathcal{B}(\mathtt{B}, \boldsymbol{b})$ and $\mathcal{S}(s,t)$ in the vector

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_G \\ \boldsymbol{\theta}_R \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} b_{11} \\ b_{21} \\ b_{12} \\ b_{22} \\ b_1 \\ b_2 \end{bmatrix} \\ \begin{bmatrix} s \\ t \end{bmatrix} \end{bmatrix} . \tag{4}$$

This model is rigorous only in the case that the scene surface is planar in a differentiable region and the intensity differences result from brightness and contrast changes only.

We now assume the intensities g_j and h_k are noisy with variances $\sigma^2_{n_j}$ and $\sigma^2_{h_k}$. These statistical properties of the noise need to be specified, *e.g.*, assuming the variance to be signal dependent, thus *e.g.* using $\sigma^2_{n_j} = \sigma^2_n(g(\boldsymbol{y}_j))$ and $\sigma^2_{m_k} = \sigma^2_m(h(\boldsymbol{z}_k))$. When using real images, we estimate this signal-dependent variance functions of the two images, see [6].

Integrating the geometry and intensity transformation we arrive at the following model, which is generative, *i.e.*, allows to simulate observed images:

$$\underline{g}(\boldsymbol{y}_j) = s^{-1}\left(f\left(\mathtt{B}\boldsymbol{y}_j + \boldsymbol{b}\right) - t\right) + \underline{n}(\boldsymbol{y}_j)\,, \quad j = 1, \ldots, J \tag{5}$$

$$\underline{h}(\boldsymbol{z}_k) = \left(sf\left(\mathtt{B}^{-1}(\boldsymbol{z}_k - \boldsymbol{b})\right) + t\right) + \underline{m}(\boldsymbol{z}_k)\,, \quad k = 1, \ldots, K\,. \tag{6}$$

1.2 Estimation

The task is to estimate the parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_G, \boldsymbol{\theta}_I)$ for the geometric and the radiometric transformation and the unknown true signal f from the observed values $g(\boldsymbol{y}_j)$ and $h(\boldsymbol{z}_k)$.

The explicit modeling in (5) and (6) allows us to write the problem as a nonlinear Gauss-Markov model with the residuals and their dispersion,

$$n_j(\boldsymbol{\theta}, f) = g_j - s^{-1}\left(f\left(\mathtt{B}\boldsymbol{y}_j + \boldsymbol{b}\right) - t\right) \quad , \quad \mathbb{D}(\underline{n}_j) = \sigma^2_{n_j}\,, \quad j = 1, ..., J \tag{7}$$

$$m_k(\boldsymbol{\theta}, f) = h_k - \left(sf\left(\mathtt{B}^{-1}(\boldsymbol{z}_k - \boldsymbol{b})\right) + t\right)\,, \quad \mathbb{D}(\underline{m}_k) = \sigma^2_{m_k}\,, \quad k = 1, ..., K\,, \tag{8}$$

for all pixels $\boldsymbol{y}_j$ of g and all pixels $\boldsymbol{z}_k$ of h falling into the common region in f. Maximum likelihood (ML) estimates $(\widehat{\boldsymbol{\theta}}, \widehat{f})$ result from minimizing the weighted sum of the residuals,

$$\Omega(\boldsymbol{\theta}, f) = \sum_j w_j n_j^2(\boldsymbol{\theta}, f) + \sum_k w_k m_k^2(\boldsymbol{\theta}, f)\,, \tag{9}$$

w.r.t. the unknown distortion parameters $\boldsymbol{\theta}$ and the unknown signal f, using proper weights

$$w_j = \frac{1}{\sigma^2_{n_j}} \quad \text{and} \quad w_k = \frac{1}{\sigma^2_{m_k}}\ . \tag{10}$$

Due to the size of f, the number of unknowns is quite large. Therefore we solve this problem by alternatively fixing one group of the parameters and solving for the other:

$$\widehat{\boldsymbol{\theta}} \mid f = \text{argmin}_{\theta} \Omega(\boldsymbol{\theta}, f)\,, \tag{11}$$

$$\widehat{f} \mid \boldsymbol{\theta} = \text{argmin}_{f} \Omega(\boldsymbol{\theta}, f)\,. \tag{12}$$

Especially, the estimated unknown function is the weighted mean of the functions g and h transformed into the coordinate system $\boldsymbol{x}$ of f, which can be calculated pixel wise:

$$\widehat{f}_i \mid \widehat{\boldsymbol{\theta}} = \frac{{}^{g}w_{f_i}\,{}^{g}f_i \;+\; {}^{h}w_{f_i}\,{}^{h}f_i}{{}^{g}w_{f_i} \;+\; {}^{h}w_{f_i}}\,, \tag{13}$$

with

$${}^{g}f_i = s \cdot g(\boldsymbol{y}_i) + t \quad \text{and} \quad {}^{h}f_i = 1/s \cdot (h(\boldsymbol{z}_i) - t) \tag{14}$$

from (3) and

$$\boldsymbol{y}_i = \mathtt{B}^{-1}(\boldsymbol{x}_i - \boldsymbol{b})) \quad \text{and} \quad \boldsymbol{z}_i = \mathtt{B}\boldsymbol{x}_i + \boldsymbol{b} \tag{15}$$

The weights are

$${}^{g}w_{f_i} = \frac{1}{s^2 \cdot V_g(g(\boldsymbol{y}_i))} \quad \text{and} \quad {}^{h}w_{f_i} = \frac{1}{V_g(h(\boldsymbol{z}_i)/s^2}\,. \tag{16}$$

Bicubic interpolation is used to transfer $g(\boldsymbol{y}_i)$ and $(\boldsymbol{z}_i)$ to $f(\boldsymbol{x}_i)$.

As a result of the ML-estimation we obtain: (1) the parameters $\widehat{\boldsymbol{\theta}}$, (2) their covariances $\Sigma_{\widehat{\theta}\widehat{\theta}}$, and (3) the variance factor

$$\widehat{\sigma}_0^2 = \frac{\Omega(\widehat{\boldsymbol{\theta}}, \widehat{f})}{R}\,, \tag{17}$$

where R is the redundancy of the system, *i.e.*, the efficient number of observations $K_g + K_h$ minus the number of unknown parameters $8 + K_f$, where we take the approximation $K_f = \sqrt{K_g K_h}$:

$$R = K_g + K_h - (8 + \sqrt{K_g K_h})\,. \tag{18}$$

If the model holds, the variance factor is Fisher distributed with $F(R, \infty)$ and should thus be close to 1. Therefore, it is reasonable to multiply the covariance matrix $\Sigma_{\widehat{\theta}\widehat{\theta}}$ with the variance factor to arrive at a realistic characterization

$$\widehat{\Sigma}_{\widehat{\theta}\widehat{\theta}} = \widehat{\sigma}_0^2 \Sigma_{\widehat{\theta}\widehat{\theta}} \tag{19}$$

of the uncertainty of the estimated parameters.

The covariance matrix $\widehat{\Sigma}_{\widehat{\psi}\widehat{\psi}}$ of the parameters in the 8-vector $\boldsymbol{\psi}$ of the geometric and radiometric affinities

$${}^{h}\mathtt{A} = \begin{bmatrix} \mathtt{A} & \boldsymbol{a} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} \psi_1 & \psi_3 & \psi_5 \\ \psi_2 & \psi_4 & \psi_6 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad {}^{h}\mathtt{R} = \begin{bmatrix} \psi_7 & \psi_8 \\ 0 & 1 \end{bmatrix} \tag{20}$$

finally is derived by variance propagation from $\mathcal{A}(\boldsymbol{\psi}) = \mathcal{B}(\boldsymbol{\theta})^2$, resulting from (1) and (2). We have

$$\mathtt{A} = \mathtt{B}^2\,, \quad \boldsymbol{a} = (\mathtt{B} + \mathtt{I}_2)\boldsymbol{b} \quad \text{and} \quad \psi_7 = \theta_7^2\,, \quad \psi_8 = (\theta_7 + 1)\theta_8\,, \tag{21}$$

with the Jacobian:

$$\mathtt{J}_{\psi\theta} = \begin{pmatrix} 2\,\theta_1 & \theta_3 & \theta_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_2 & \theta_1 + \theta_4 & 0 & \theta_2 & 0 & 0 & 0 & 0 & 0 \\ \theta_3 & 0 & \theta_1 + \theta_4 & \theta_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & \theta_3 & \theta_2 & 2\,\theta_4 & 0 & 0 & 0 & 0 & 0 \\ \theta_5 & 0 & \theta_6 & 0 & \theta_1 + 1 & \theta_3 & 0 & 0 & 0 \\ 0 & \theta_5 & 0 & \theta_6 & \theta_2 & \theta_4 + 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2\,\theta_7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \theta_8 & \theta_7 + 1 & 0 \end{pmatrix}. \tag{22}$$

Empirical tests with simulated data confirm the desired properties: (1) Exchanging the two images g and h leads to the inverse transformation $\mathcal{A}^{-1}$, (2) the covariance matrix derived from samples with different noise do not show significant deviations from the theoretical covariance matrix, and (3) the mean variance factor is not much larger than 1, usually by 20% to 40%. The deviations from 1 are significant and can be explained by the approximations resulting from the bicubic interpolations of the same function based on different grids. Hence the internal quality measures can be used for self-diagnosis.

We now show the potential of the refinement of the affinitiy using LSM, namely: the expected precision of the affinity for ideal cases. This is based on the part of the normal equation matrix $\mathtt{N}$ related to the 6 parameters of the geometric affinity: $\mathtt{N} = \sigma_n^{-2} \sum_{ij} \nabla f_\theta(i,j) \nabla f_\theta^\mathsf{T}(i,j)$, where the sum is over all pixels in an $N \times N$ window. If we assume the distortion is zero and f is known, then the gradient is $\nabla f = [x f_x, y f_x, x f_y, y f_y, f_x, f_y]$. Observe, the 2×2 matrix referring to the translation parameters is proportional to the structure tensor of the patch. We now assume that the gradients in the window have the same variance $\sigma_{f'}^2$ and are mutually uncorrelated. Then the normal equation matrix will be diagonal leading to the covariance matrix

$$\Sigma_{\alpha\alpha} = \begin{bmatrix} \sigma_a^2 \mathtt{I}_4 & \mathtt{0} \\ \mathtt{0} & \sigma_p^2 \mathtt{I}_3 \end{bmatrix} \quad \text{with} \quad \sigma_a = \frac{\sqrt{12}}{N^2} \frac{\sigma_n}{\sigma_{f'}} \quad \text{and} \quad \sigma_p = \frac{1}{N} \frac{\sigma_n}{\sigma_{f'}}\,. \tag{23}$$

Fig. 2 shows the predicted standard deviations for real image patches having different sizes, depending on the LOWE-scale s using $M = 7s$. In spite of the variation of the texture within the individual patches, the theoretical relation to the window size is visible. Deviations for large scales result from the fact that sometimes the interior of such patches is not highly textured.

2 Solvers using ACs – Semi-calibrated case

In this section, the solver for AC-based relative pose and focal length estimation is discussed.

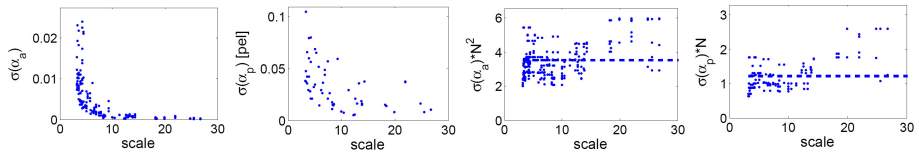


Fig. 2: Predicted standard deviations of affine parameters as a function of the LOWE-scale of the affine correspondence. **1 and 2:** Non-normalized values $\sigma_{\alpha_i}, i = 1, ..., 4$ are unit less, the values $\sigma_{\alpha_i}, i = 5, 6$ have unit [pixel]. Observe, subpixel accuracy can be reached and the affinity parameters have a standard deviation below 1%, except for very small scales. **3 and 4:** Normalized standard deviations $N^2\,\sigma_{\alpha_i}, i = 1, ..., 4$ and $N\sigma_{\alpha_i}, i = 5, 6$ would be constant if the texture in all windows had the same gradient variance. The deviations from a constant reflect the variation of the texture in the images. Observe, the ratio between the mean values for the normalized standard deviations of the affine and the shift parameters is 3.7/1.3=2.77 (see the dashed lines in 3 and 4). It is in fair coherence with the theoretical value $\sqrt{12} = 3.46$

Essential matrix and focal length from 2ACs: The problem of estimating the relative pose and a common focal length of two semi-calibrated cameras, *i.e.*, estimating the unknown essential matrix $\mathtt{E} \in \mathbb{R}^{3\times3}$ and focal length $f \in \mathbb{R}$, has six degrees of freedom (three for rotation, two for translation and one for focal length) and there exists several well-known hidden variable or Gröbner basis 6PCs solvers [**?**,9]. This problem has recently been solved from 2ACs [1]. Each AC gives three linear constraints [3]. One approach to solve for $\mathtt{E}$ and f from 2ACs is to apply the solver of [1]. The 2ACs solver [1] as well as the hidden variable and Gröbner basis 6PCs solvers [**?**,9] are based on ten well-known constraints on the essential matrix $\mathtt{E}$, *i.e.*, the singularity constraint $\det(\mathtt{E}) = 0$ and the trace constraint $2\mathtt{E}\mathtt{E}^\top\mathtt{E} - \mathrm{tr}(\mathtt{E}\mathtt{E}^\top)\mathtt{E} = \mathtt{0}$. These solvers are therefore solving a system of ten polynomial equations in three unknowns.

Recently, however, a more efficient solver for the 6PC problem was proposed in [10]. This solver is based on the fact that for a fundamental matrix of the form $\mathtt{F} = \mathtt{K}^{-\top}\mathtt{E}\mathtt{K}^{-1}$, where $\mathtt{K} = \mathrm{Diag}([f, f, 1])$ is the calibration matrix with focal length f, two constraints can be derived. These constraints can by obtained by eliminating the focal length from the singularity and the trace constraint. The first constraint corresponds to the singularity constraint and the second one is a fifth degree polynomial in the elements of $\mathtt{F}$ (see Eq 18 in [10]). The final solver, after calculating the three dimensional null-space using six equations from the epipolar constraint, then solves these two equations in two unknowns (unknowns from the parameterization of $\mathtt{F}$ as a linear combination of basis null-space vectors). Note that in the solver the scale of $\mathtt{F}$ is fixed by fixing one element in the linear combination of the null-space vectors. However, for the uncertainty propagation the scale will be fixed by applying a different constraint on $\mathtt{F}$, *i.e.*, the constraint $||\mathrm{vec}(\mathtt{F})|| - 1 = 0$. The parameterization from [10], with two constraints instead of ten from the singularity and the trace constraints [**?**,9], not only leads

to a simpler solver but will also be useful for fast uncertainty propagation as we will discuss in Sec. 3.4.

The formulation with two constraints on $\mathtt{F}$ can be straightforwardly adapted for the 2ACs solver. We propose to use the solver from [10] by first calculating the null-space using the linear system which 2 ACs imply. The rest of the solver remains unchanged and can be similarly applied as when using point correspondences.

3 Uncertainty Calculation

All geometric problems we address in the paper are based on a set of constraints $\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta}) = \mathbf{0}$ between some observations $\boldsymbol{y}$, and some parameters $\boldsymbol{\theta}$. The classical variance propagation for implicit functions leads to

$$\Sigma_{\theta\theta} = \mathtt{B}^{-1}\mathtt{A}\Sigma_{yy}\mathtt{A}^{\mathsf{T}}\mathtt{B}^{-\mathsf{T}} \quad \text{with} \quad \mathtt{A} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}} \quad \text{and} \quad \mathtt{B} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\theta}}\,, \tag{24}$$

see [7, Sect. 2.7.5]. In our context, the covariance matrix Σ_{yy} refers to the input measurements (*e.g.*, keypoints coordinates or affinity correspondences), the covariance matrix $\Sigma_{\theta\theta}$ refers to the model parameters.

For minimal problems, the number of constraints $\boldsymbol{g}$ is usually smaller than the number of parameters of the model and hence the matrix $\mathtt{B}$ cannot be inverted. Therefore, we propose to redefine the implicit function by adding constraints $\boldsymbol{h}(\boldsymbol{\theta}) = \mathbf{0}$ between the model parameters only, *i.e.*, we use the extended implicit function with their Jacobians

$$\begin{bmatrix} \boldsymbol{g}(\boldsymbol{y}, \boldsymbol{\theta}) \\ \boldsymbol{h}(\boldsymbol{\theta}) \end{bmatrix} = \mathbf{0} \quad \text{with} \quad \mathtt{A} = \begin{bmatrix} \partial \boldsymbol{g}/\partial \boldsymbol{y} \\ \mathtt{0}^{\mathsf{T}} \end{bmatrix}, \quad \mathtt{B} = \begin{bmatrix} \partial \boldsymbol{g}/\partial \boldsymbol{\theta} \\ \partial \boldsymbol{h}/\partial \boldsymbol{\theta} \end{bmatrix}. \tag{25}$$

This way we have exactly the same number of constraints as the parameters, and – except for critical geometric configurations – the matrix $\mathtt{B}$ is regular. The following paragraphs list the minimal set of constraints $\boldsymbol{g}$, $\boldsymbol{h}$ used by individual solvers in our paper.

3.1 Homography estimation

The homography matrix has eight degrees of freedom and is defined by the nine parameters $\boldsymbol{\theta}$, *i.e.*, the elements of the matrix $\mathtt{H} \in \mathbb{R}^{3\times3}$. In the case of PCs, we used eight constraints $\boldsymbol{g}$ from 4 PCs following [8] (the DLT algorithm for homography estimation) and one constraint $h(\mathtt{H}) = ||\mathrm{vec}(\mathtt{H})|| - 1 = 0$, which avoids the trivial all-zeros solution. Observe, the this norm constraint influences both, the scale of $\mathtt{H}$ and the scale of its covariance matrix.

Assuming two ACs, one can select the subset of constraints used for the uncertainty propagation. We defined $\boldsymbol{g}$ as four constraints from two point correspondences [2] (Eqn. 1 in [2]), and four constraints from one affinity matrix [2] (Eqn. 4 in [2]). The constraint $h(\mathtt{H}) = ||\mathrm{vec}(\mathtt{H})|| - 1 = 0$ is the same as for PCs.

3.2 Fundamental matrix estimation

The fundamental matrix has seven degrees of freedom and is defined by the nine parameters of the matrix $\mathtt{F} \in \mathbb{R}^{3\times3}$. In the case of PCs, we used seven constraints $\boldsymbol{g}_i = \boldsymbol{y}_i^\top \mathtt{F} \boldsymbol{z}_i = 0$ where $i \in \{1, \dots, 7\}$ and two constraints $\boldsymbol{h} = [h_1, h_2]^\mathsf{T} = \mathbf{0}$, where $h_1(\mathtt{F}) = \det(\mathtt{F}) - 1 = 0$ and $h_2(\mathtt{F}) = ||\text{vec}(\mathtt{F})|| - 1 = 0$.

The constraints for 3 ACs are composed of (1) three point constraints following [8] of the form

$$c_{p_i} := \boldsymbol{z}_i^\mathsf{T} \, \mathtt{F} \, \boldsymbol{y}_i = 0 \,, \quad i = 1, 2, 3 \,, \tag{26}$$

(2) three pairs of affine constraints [1], (Eqn. 8 in [1]) of the form

$$\underset{2\times1}{\boldsymbol{c}_{a_i}} := [\mathtt{I}_2 \mid \mathbf{0}] \, \mathtt{F} \, \boldsymbol{y}_i + [\mathtt{A}_i^{-\mathsf{T}} \mid \mathbf{0}] \, \mathtt{F}^\mathsf{T} \, \boldsymbol{z}_i = \mathbf{0} \,, \quad i = 1, 2, 3 \,, \tag{27}$$

and (3) the two constraints $\boldsymbol{h} = [h_1, h_2]^\mathsf{T} = \mathbf{0}$. The propagation function uses the three point constraints, *i.e.*, $c_{p_i} = 0$ $(i = 1, 2, 3)$, the first two pairs $\boldsymbol{c}_{a_i} = \mathbf{0}$ $(i = 1, 2)$ of the three pairs of affine constraints, and the two constraints $\boldsymbol{h} = \mathbf{0}$ on $\mathtt{F}$.

3.3 Essential matrix estimation

We list two ways how to propagate the uncertainty to the parameters of the essential matrix. The essential matrix $\mathtt{E} \in \mathbb{R}^{3\times3}$ has nine variables and five degrees of freedom. Assuming point correspondences, a straightforward solution is to employ five points constraints $\boldsymbol{g}_i = \boldsymbol{y}_i^\top \mathtt{E} \boldsymbol{z}_i = 0$, where $i \in \{1, \dots, 5\}$ and four constraints $\boldsymbol{h} = [h_1, h_2, h_3, h_4]^\mathsf{T}$ on $\text{vec}(\mathtt{E})$. The constraints $h_1(\mathtt{E}) = \det(\mathtt{E}) = 0$ and $h_2(\mathtt{E}) = ||\text{vec}(\mathtt{E})||^2 - 2 = 0$ correspond to those for the fundamental matrix. The essential matrix has also nine trace constraints $\mathtt{C} := 2\mathtt{E}\mathtt{E}^\top\mathtt{E} - \text{tr}(\mathtt{E}\mathtt{E}^\top)\mathtt{E} = \mathtt{0}$. Generally, we can select any two of this nine constraints such that the constraints in $\boldsymbol{h}$ are independent, except for certain cases, such as $\mathtt{E} = [1; 0; 0]_\times$, where the Jacobian $\partial\text{vec}(\mathtt{C})/\partial\text{vec}(\mathtt{E})$ has 4 zero rows, not allowing us to use the corresponding constraints, and one is proportional to the Jacobian of the determinant constraint.

The propagation can also be obtained by using a minimal set of parameters, *e.g.*, a unit translation vector $\boldsymbol{b} = [b_1, b_2, b_3]^\top$ and the Euler vector parametrization of the rotation $\mathtt{R}(\boldsymbol{r}) \in \text{SO}(3)$. For this parametrization, we assume the essential matrix has the form $\mathtt{E} = [\boldsymbol{b}]_x \mathtt{R}(\boldsymbol{r})$, where

$$[\boldsymbol{b}]_\times = \begin{bmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{bmatrix} \quad , \qquad \alpha = \sqrt{\boldsymbol{r}^\top \boldsymbol{r}} \ , \tag{28}$$

$$\mathtt{R}(\boldsymbol{r}) = \mathtt{I}_3 + (1 - \cos\alpha) \, [\boldsymbol{r}]_\times^2 + \sin\alpha \, [\boldsymbol{r}]_\times \ . \tag{29}$$

Here, the matrix $\mathtt{I}_3 \in \mathbb{R}^{3x3}$ is the identity matrix. This representation leads to five points constraints and only one constraint $h(\boldsymbol{b}) = ||\boldsymbol{b}|| - 1 = 0$. We used this second representation for our experiments.

The minimal problem using two AC has three constraints for each point, *i.e.*, $\boldsymbol{g}_i = \boldsymbol{y}_i^\top \mathtt{E} \boldsymbol{z}_i = 0$, and constraints provided by Eqn. 9 and 10 in [1]. We used the reduced set of parameters $\boldsymbol{b}$, $\boldsymbol{r}$ with the constraint $h(\boldsymbol{b}) = ||\boldsymbol{b}|| - 1 = 0$ and five of six constraints on the two ACs, *i.e.*, we suppress one equation of the form of Eq. 10 in [1] to have the same number of constraints as parameters.

3.4 Essential matrix + focal length estimation

The essential matrix with focal length has six degrees of freedom and nine parameters as it can be described as $\mathtt{F} = \mathtt{K}^{-\top}\mathtt{E}\mathtt{K}^{-1}$, where $\mathtt{K} = \text{Diag}([f, f, 1])$ is the calibration matrix with focal length f. We assume six point or two affine correspondences as input, which lead to the same constraints as for the essential matrix, *i.e.*, six constraints $g(\boldsymbol{\theta})$. Further, we used the constraints $\boldsymbol{h} = [h_1, h_2, h_3]^\mathsf{T}$ where h_1, h_2 are the same constraints as for the fundamental matrix and h_3 corresponds to Eqn. 18 from [10].

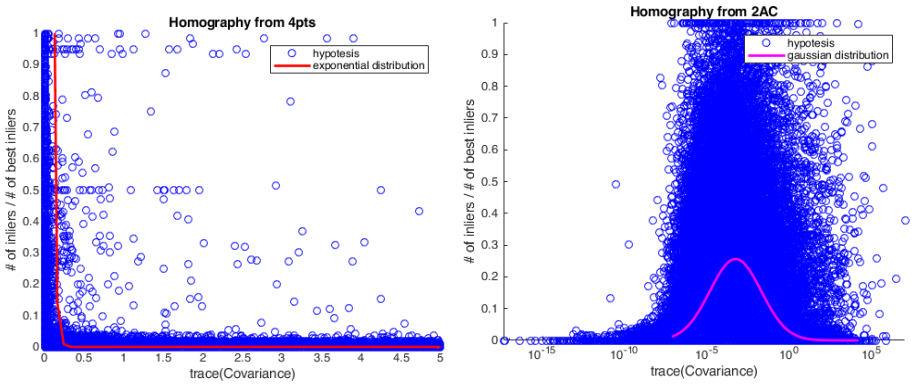
4 Preemptive Verification

We used the a-priori determined parameters of the inlier ratio distributions (*c.f.* Fig. 3), to measure the probability of having a *good model* based on its uncertainty. For determining the parameters of the distributions, we saved the uncertainty (measured by the trace of the covariance matrix) of all models generated in the RANSAC loop, on all scenes from all datasets, and, also, their number of inliers divided by the number of inliers of the best model found on that particular image pair, *i.e.*, we saved their inlier ratios. We found that calculating the parameters of the distributions using only those models which lead to an inlier ratio of, at least, 0.95 works well for for all problems considered in the paper.

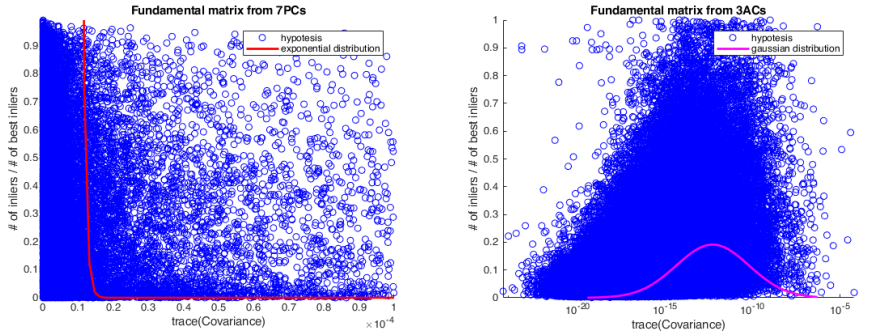
In our experiments, we fit the exponential distribution parameter $\hat{\lambda} = (n - 2)/\sum_{i=1}^{n}(t_i)$, where n is the number of good models θ_i and t_i is the trace of covariance matrices, *i.e.*, $t_i = \text{tr}(\Sigma_{\theta_i\theta_i})$. For solvers using affine correspondences, the distributions are described by a log-normal distribution, *i.e.*, by mean $\mu = 1/n \sum_{i=1}^{n} \log_{10}(t_i)$ and variance $\sigma^2 = 1/(n-1)\sum_{i=1}^{n}(\log_{10}(t_i) - \mu)^2)$.

5 Local Optimization

The cumulative distribution functions of the wall-clock times and $\log_{10}$ iteration numbers using different local optimizations for affine correspondence (AC) and point-based (PC) solvers are shown in Fig. 4. For homography and fundamental matrix estimation, the proposed combined preemptive model verification strategy was used. For essential matrices, SPRT [5] was applied. It can be seen that Graph-Cut RANSAC [4] (GC-RANSAC) is the fastest and leads to the fewest RANSAC iterations on all investigated problems. Also, methods using ACs are significantly faster than point-based ones.

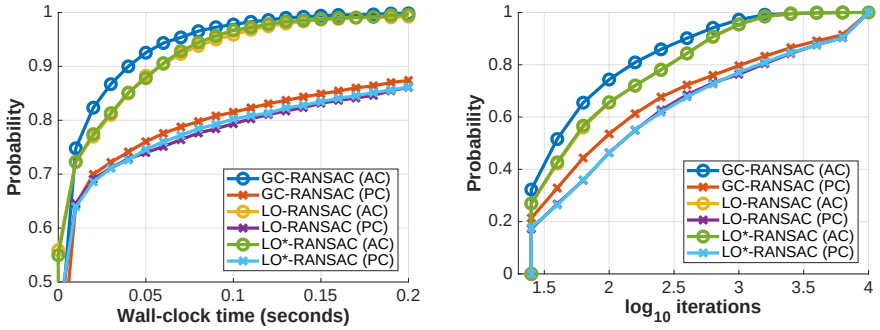


(a) Homography estimation

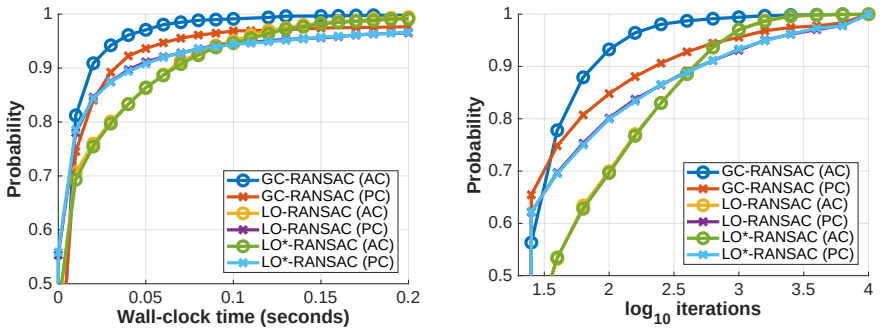


(b) Fundamental matrix estimation

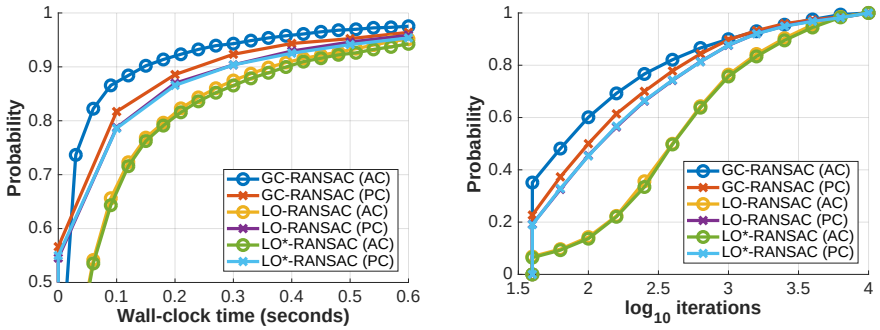
Fig. 3: The inlier ratio (vertical) of homographies and fundamental matrices from all tested real scenes as a function of the trace of their covariance matrices (horizontal). This shows that, for points, uncertain models (on right) generate small numbers of inliers. For affine correspondences, the traces have a log-normal distribution.



(a) Homography estimation



(b) Fundamental matrix estimation



(c) Essential matrix estimation

Fig. 4: The cumulative distribution functions of the wall-clock times and $\log_{10}$ iteration numbers using different local optimizations for affine correspondence (AC) and point-based (PC) solvers.

References

1. 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)
2. Barath, D., Hajder, L.: A theory of point-wise homography estimation. Pattern Recognition Letters **94**, 7–14 (2017)
3. Barath, D., Hajder, L.: Efficient recovery of essential matrix from two affine correspondences. IEEE Transactions on Image Processing **27**(11), 5328–5337 (2018)
4. Barath, D., Matas, J.: Graph-cut RANSAC. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6733–6741 (2018)
5. Chum, O., Matas, J.: Optimal randomized RANSAC. IEEE Transactions on Pattern Analysis and Machine Intelligence **30**(8), 1472–1482 (2008)
6. 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
7. Förstner, W., Wrobel, B.P.: Photogrammetric Computer Vision. Springer (2016), http://www.ipb.uni-bonn.de/book-pcv/
8. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
9. Hartley, R., Li, H.: An efficient hidden variable approach to minimal-case camera motion estimation. IEEE transactions on pattern analysis and machine intelligence **34**(12), 2303–2314 (2012)
10. Kukelova, Z., Kileel, J., Sturmfels, B., Pajdla, T.: A clever elimination strategy for efficient minimal solvers. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4912–4921 (2017)