Supplementary Material:
Relative Pose from SIFT Features

Daniel Barath[1] and Zuzana Kukelova[2]

[1] ETH Zurich, Computer Vision and Geometry Group, Switzerland
danielbela.barath@inf.ethz.ch
[2] Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague
kukelzuz@fel.cvut.cz

1 Additional Experiments

While we propose a principled way of estimating the relative pose from SIFT correspondence, there are possible approximations that are often used in the literature. When having SIFT orientations and scales, affine correspondence-based solvers are not applicable directly. However, they can be approximated in each image as $\mathbf{A}_i = \mathbf{R}_i \mathbf{S}_i$ where $\mathbf{R}_i \in \mathrm{SO}(2)$ rotates by the SIFT orientation and $\mathbf{S}_i \in \mathbb{R}^{2 \times 2}$ scales by the feature size, where $i \in [1, 2]$ is the image index. Another often used approximation is converting the SIFT features to point correspondences [6].

We ran fundamental matrix estimation on KITTI and PhotoTourism using the TIP solver from [6], and the affine solver from [2] on the generated point correspondences and approximated affine features, respectively. Similarly as for the experiments in the main paper, we ran the methods on KITTI using multiple frame differences, *i.e.*, 1, 2 and 4. The average rotation ($\epsilon_{\mathbf{R}}$) and translation errors ($\epsilon_{\mathbf{t}}$), the number of inliers, run-times in milliseconds and number of iterations are shown in Table 1.

	PhotoTourism					**KITTI**				
	$\epsilon_{\mathbf{R}}$	$\epsilon_{\mathbf{t}}$	# inliers	time (ms)	# iters	$\epsilon_{\mathbf{R}}$	$\epsilon_{\mathbf{t}}$	# inliers	time (ms)	# iters
Proposed	**2.1**	**6.7**	267	**48.8**	**4189**	**2.7**	**2.2**	1676	67.3	304
SIFT-to-1AC	5.9	12.5	239	236.4	4845	**2.7**	2.3	1676	**56.9**	**152**
SIFT-to-2PC	2.3	7.5	262	208.7	4264	2.8	3.1	1652	93.7	363
PC-based	2.3	7.8	**268**	127.3	7145	**2.7**	2.3	**1677**	154.4	1860

Table 1. The average rotation ($\epsilon_{\mathbf{R}}$) and translation errors ($\epsilon_{\mathbf{t}}$) in degrees, the number of inliers (# inliers), run-times in milliseconds and number of iterations (# iters) on the KITTI and PhotoTourism datasets. The compared methods are the proposed one, the method approximating affine correspondences (ACs) from the SIFT features and applying the AC-based solver from [2], the method converting the SIFT features to point correspondences (PCs) by [6], and the standard 7 PC-based solver.

On KITTI, the differences are small due to the dataset being easy. Still, the proposed solver leads to the best results and it is the second fastest by being marginally slower than [2] that uses fewer matches. On PhotoTourism, as it is a challenging dataset with diverse camera motions, the SIFT-to-1AC [2] and SIFT-to-2PC [6] approximations do not work as well as on KITTI. The AC-based method fails completely on PhotoTourism. The solver from [6] is marginally better than the PC-based approach but it is significantly slower. The proposed solver is the fastest by a large margin while also being the most accurate method.

2 Constraints Relating Elements of A and SIFT Parameters

In the main paper, in Section 3.1, we discussed constraints relating elements of the matrix $\mathbf{A}$ and SIFT parameters. The first set of constraints was derived from the decomposition of $\mathbf{A}$ as the multiplication of the Jacobians of the projection functions w.r.t. the image directions in the two images. These equations have, after simplification, the following form (Eq. (11) in the main paper):

$$\begin{aligned} a_1 &= c_2 c_1 q_u - c_2 s_1 w + s_2 s_1 q_v, \\ a_2 &= c_2 s_1 q_u + c_2 c_1 w - s_2 c_1 q_v, \\ a_3 &= s_2 c_1 q_u - s_2 s_1 w - c_2 s_1 q_v, \\ a_4 &= s_2 s_1 q_u + s_2 c_1 w + c_2 c_1 q_v, \end{aligned} \tag{1}$$

where the unknowns are the affine parameters a_1, a_2, a_3, a_4, scales q_u, q_v and shear w. Angles α_1 and α_2 are known from the SIFT features and $c_i = \cos(\alpha_i)$ and $s_i = \sin(\alpha_i)$.

In addition to these constraints, there are two more constraints relating SIFT parameters and $\mathbf{A}$. First, the uniform scales of the SIFT features are proportional to the area of the underlying image region and, therefore, the scale change provides constraint (Eq. (12) in the main paper):

$$\det \mathbf{A} = \det \left(\mathbf{R}_2 \mathbf{U} \mathbf{R}_1^{\mathrm{T}} \right) = \det \mathbf{U} = q_u q_v = \frac{q_2^2}{q_1^2}, \tag{2}$$

where q_1 and q_2 are the SIFT scales in the two images. Second, the oriented circles centered on the point correspondence provide an additional constraint of the form (Eq. (13) in the main paper):

$$q_1 \mathbf{A} \begin{bmatrix} \cos(\alpha_1) \\ \sin(\alpha_1) \end{bmatrix} = q_2 \begin{bmatrix} \cos(\alpha_2) \\ \sin(\alpha_2) \end{bmatrix}. \tag{3}$$

The constraints (2) and (3) can be rewritten as

$$a_2 a_3 - a_1 a_4 + q^2 = 0, \tag{4}$$
$$a_3 c_1 + a_4 s_1 - s_2 q = 0, \tag{5}$$
$$a_1 c_1 + a_2 s_1 - c_2 q = 0, \tag{6}$$

where $q = \frac{q_2}{q_1}$.

Here, we show that constraints (4)-(6) constitute all constraints that relate the elements of $\mathbf{A}$ and the measured orientations α_i and scales $q_i,\ i = 1,2$ of the features in the first and second image. In other words, equations (1) are not adding any additional constraints relating the elements of $\mathbf{A}$ and the measured orientations α_i and scales $q_i,\ i = 1,2$ to constraints (4)-(6).

To prove this, we first define the ideal I [3] generated by polynomials (1),(2),(3) and trigonometric identities $c_i^2 + s_i^2 = 1$ for $i \in \{1,2\}$. To transform (2) to a polynomial equation, we substitute $q = \frac{q_2}{q_1}$ and add a constraint $q_1 q - q_2 = 0$ to the ideal. Moreover, we ensure $q_1 \neq 0$ by saturating the ideal with q_1[3]. Note that here we consider all elements of these polynomials, including q, q_i, c_i and s_i, as unknowns. Then we compute the generators of the elimination ideal $I_1 = I \cap \mathbb{C}[a_1, a_2, a_3, a_4, q, s_1, c_1, s_2, c_2]$ [3]. The generators of I_1 do not contain q_u, q_v and w and, instead, of q_1 and q_2 they directly contain $q = \frac{q_2}{q_1}$. These generators are exactly equations (4)-(6) together with trigonometric identities $c_i^2 + s_i^2 = 1$ for $i \in \{1,2\}$. This means that constraints (1) do not add any additional information and the elimination ideal is directly generated by (2) and (3), or equivalently by (4)-(6)[4]. Generators (4)-(6) can be computed using a computer algebra system, *e.g.*, `Macaulay2` [4]. The input code for `Macaulay2` is included in the supplementary material as a separate constraints_A_SIFT.m2 file.

On the other hand, the constraints (1) and (2) that were used for derivations in [1], are not covering all constraints that relate the elements of $\mathbf{A}$ and the measured orientations α_i and scales $q_i,\ i = 1,2$. This can be easily proved. By eliminating q_u, q_v and w from the ideal generated by (1), (2) and trigonometric identities $c_i^2 + s_i^2 = 1$ for $i \in \{1,2\}$, using the elimination ideal technique [5][5], we obtain two generators of the elimination ideal. One generator is directly (4), *i.e.* the constraint (2), and the second one has the form

$$c_1 s_2 a_1 + s_1 s_2 a_2 - c_1 c_2 a_3 - c_2 s_1 a_4 = 0. \tag{7}$$

The constraints (4) and (7) are constraints proposed in [1]

The constraint (7) is a linear combination of constraints (5) and (6) with coefficients $-c_2$ and s_2. This means that if (5) and (6) vanish also (7) vanishes. However, the opposite is not true. If (7) vanishes the constraints (5) and (6) do not need to vanish. This means that the constraints derived in [1], *i.e.*, constraints (4) and (7), do not cover all constraints that relate the elements of $\mathbf{A}$ and the measured orientations α_i and scales $q_i,\ i = 1,2$. For deriving all constraints, the constraint (3) is important.

[3] Note that geometrically $q_1 \neq 0$ and $q_2 \neq 0$, but algebraically it is sufficient to remove solutions $q_1 = 0$ by saturating the ideal with q_1. Saturating the ideal with both q_1 and q_2 will not affect the solutions.

[4] Note that for correct derivations, all steps including saturation and adding trigonometric identities to the ideal are important.

[5] Note, that here we again substitute $q = \frac{q_2}{q_1}$, add the constraint $q_1 q - q_2 = 0$ to the ideal, and saturate the ideal with q_1.

3 SIFT Epipolar Constraint

In the main paper, in Section 3.2., we derived a new constraint relating epipolar geometry and the measured orientations α_i and scales $q_i,\ i = 1, 2$ of covariant features in the first and second image. For this purpose, we used elimination ideal technique [5]. The input code for `Macaulay2` used to compute this constraint, *i.e.*, the generator of the elimination ideal J_1 is provided as a separate `Macaulay2` constraint_F_SIFT.m2 file.

References

1. Barath, D., Kukelova, Z.: Homography from two orientation-and scale-covariant features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1091–1099 (2019)
2. Barath, D., Polic, M., Förstner, W., Sattler, T., Pajdla, T., Kukelova, Z.: Making affine correspondences work in camera geometry computation. In: European Conference on Computer Vision. pp. 723–740. Springer (2020)
3. Cox, D., Little, J., O'Shea, D.: Using Algebraic Geometry. Springer-Verlag New York, 2nd edn. (2005), http://www.cs.amherst.edu/ dac/uag.html
4. Grayson, D., Stillman, M.: Macaulay2, a software system for research in algebraic geometry, available at `www.math.uiuc.edu/Macaulay2/`
5. Kukelova, Z., Kileel, J., Sturmfels, B., Pajdla, T.: A clever elimination strategy for efficient minimal solvers. In: Conference on Computer Vision and Pattern Recognition (2017), http://arxiv.org/abs/1703.05289
6. Riggi, F., Toews, M., Arbel, T.: Fundamental matrix estimation via tip-transfer of invariant parameters. In: International Conference on Pattern Recognition. vol. 2, pp. 21–24. IEEE (2006)