Homography-based Egomotion Estimation Using Gravity and SIFT Features
-
Supplementary material

Yaqing Ding[1], Daniel Barath[2,3], and Zuzana Kukelova[3]

[1] Nanjing University of Science and Technology, China
dingyaqing@njust.edu.cn
[2] Machine Perception Research Laboratory, SZTAKI in Budapest
barath.daniel@sztaki.mta.hu
[3] VRG, Faculty of Electrical Engineering, Czech Technical University in Prague
kukelova@cmp.felk.cvut.cz

This supplementary material provides the following information: Section 1 describes a 2ORB solver for correspondences of points/features originating from a general plane. Section 2 provides additional synthetic experiments. Section 3 provides a proof that the rotation estimation is independent of the scale for the 1SIFT solver.

1 Points on a General Plane

We consider the case where the normal of the plane is unknown and has arbitrary orientation. It is a 6-DOF problem with respect to $\{\theta, t_x, t_y, t_z, n_x, n_y, n_z\}$, where θ is the rotation around the vertical axis; t_x, t_y, t_z are the coordinates of the translation; and n_x, n_y, n_z define the unit-length surface normal, i.e., $n_x^2 + n_y^2 + n_z^2 = 1$. We choose to use two orientation-covariant features, e.g., ORB [1], which provide six linear constraints. These constraints can be written in matrix form as

$$\mathtt{M}\mathtt{g} = \mathtt{0} \ , \tag{1}$$

where $\mathtt{M}$ is a 6×9 coefficient matrix and the vector $\mathtt{g}$ contains the elements of $\mathtt{G}$. The vector $\mathtt{g}$ can be written as a linear combination of the three basis vectors from the 3-dimensional null space of the matrix $\mathtt{M}$ as

$$\mathtt{g} = x_1\mathtt{g}_a + x_2\mathtt{g}_b + x_3\mathtt{g}_c \ , \tag{2}$$

where we can fix $x_3 = 1$. Substituting (2) into $\widehat{\mathtt{H}}_y = \mathtt{R}_2\mathtt{K}_2^{-1}\mathtt{G}\mathtt{K}_1\mathtt{R}_1^\top$, the Euclidean homography matrix $\widehat{\mathtt{H}}_y$ is parameterized using two unknowns $\{x_1, x_2\}$. The matrix $\mathtt{B}$ is defined as

$$\begin{aligned} \mathtt{B} &= \widehat{\mathtt{H}}_y - \lambda\mathtt{R}_y \\ &= \lambda\mathtt{t}\mathtt{n}^\top \ . \end{aligned} \tag{3}$$

As shown in [2], the matrix $\mathtt{B}$ should satisfy $\text{rank}(\mathtt{B}) = 1$. Matrix $\mathtt{B}$ is formulated as

$$\mathtt{B} = \begin{bmatrix} h_1 - \alpha & h_2 & h_3 - \beta \\ h_4 & h_5 - \lambda & h_6 \\ h_7 + \beta & h_8 & h_9 - \alpha \end{bmatrix} , \tag{4}$$

where $\alpha = \lambda \cos\theta$, $\beta = \lambda \sin\theta$, and θ is the yaw angle. Since $\text{rank}(\mathtt{B}) = 1$, each of the 2×2 submatrices of $\mathtt{B}$ must have zero determinant and we obtain nine polynomial equations. Based on $\alpha^2 + \beta^2 = \lambda^2$, we can eliminate α, β from the nine equations and obtain five constraints on the elements of $\widehat{\mathtt{H}}_y$

$$\begin{aligned} (h_1 - h_9)(h_5 - \lambda) - h_2 h_4 + h_6 h_8 = 0 \ , \\ (h_3 + h_7)(h_5 - \lambda) - h_2 h_6 - h_4 h_8 = 0 \ , \\ \lambda^2(h_2^2 + h_8^2) - (h_1 h_8 - h_2 h_7)^2 - (h_3 h_8 - h_2 h_9)^2 = 0 \ , \\ \lambda^2(h_4^2 + h_6^2) - (h_1 h_6 - h_3 h_4)^2 - (h_4 h_9 - h_6 h_7)^2 = 0 \ , \\ (h_5 - \lambda)(h_7^2 + h_9^2 - \lambda^2) + (h_3 - h_7) h_4 h_8 - (h_1 + h_9) h_6 h_8 = 0 \ . \end{aligned} \tag{5}$$

Substituting the formulation of $\widehat{\mathtt{H}}_y$ into (5) results in five polynomials w.r.t three unknowns $\{x_1, x_2, \lambda\}$. Finally, using an automatic generator for the polynomial equations, e.g., [3, 4], we obtain a template of size 24×34 for Gaussian-Jordan elimination.

2 Additional Synthetic Experiments

In this section, we show the results from the synthetic experiments under sideways motion. Fig. 1 reports the rotation and translation errors for points on the ground plane. The top row shows the performance under image noise with different standard deviations. The bottom row shows the performance with increased directional noise and constant image noise of 2 pixel standard deviation. Fig. 2 reports the rotation and translation errors for points on a vertical plane. The top row shows the performance under image noise with different standard deviations. The bottom row shows the performance with increased directional noise and constant image noise of 2 pixel standard deviation. Fig. 3 reports the rotation and translation errors for points on a general plane. The top row shows the performance under image noise with different standard deviations. The bottom row shows the performance with increased directional noise and constant image noise of 2 pixel standard deviation.

3 Proof that the Rotation Estimation is Independent of the Feature Scale for the 1SIFT Solver

Assume that the y-axes of the cameras have been aligned with the gravity direction. We have 1 SIFT correspondence: $\mathtt{m}_1 = [u_1, v_1, \varphi_1, q_1] \leftrightarrow \mathtt{m}_2 = [u_2, v_2, \varphi_2, q_2]$.

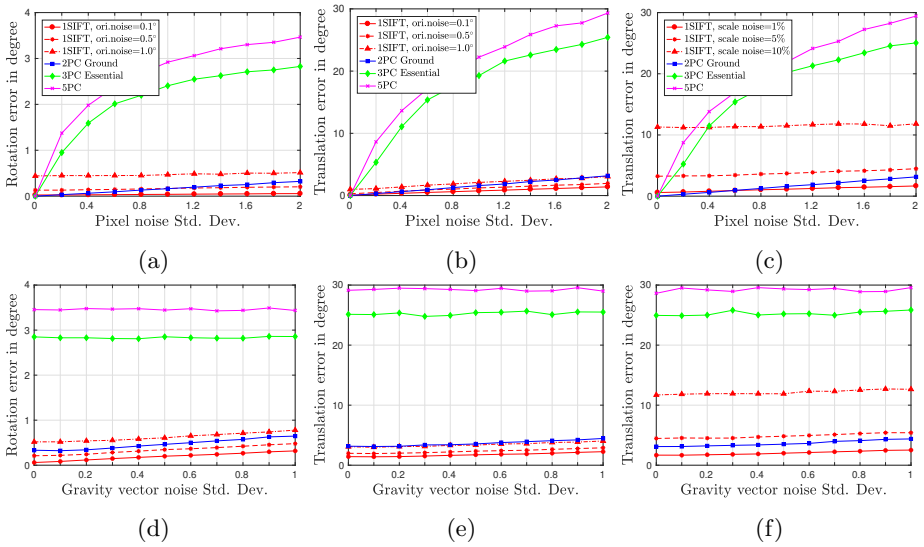

Fig. 1: Comparing the proposed 1SIFT solver with the 2PT solver of Saurer et al. [5] when points are on the ground (2PC), the three-point essential matrix-based solver of Fraundorfer et al. (3PC Essential) [6] and the five-point solver of Nister (5PC) [7]. The cameras undergo sideways motions. **Top**: Performance under increasing pixel noise. (a–b) Rotation and translation errors with additional noise added to SIFT orientations. (c) Translation error with additional noise added to SIFT scales. **Bottom**: Performance under increased directional noise and constant image noise of 2 pixel standard deviation. (d–e) Rotation and translation errors with additional noise added to SIFT orientations. (f) Translation error with additional noise added to SIFT scales.

A horizontal plane-induced Euclidean homography $\mathtt{H_y}$ should obey

$$h_4 = 0, \quad h_6 = 0, \quad h_1 - h_9 = 0, \quad h_3 + h_7 = 0 \ . \tag{6}$$

In this case, we can write

$$\mathtt{H_y} = \begin{bmatrix} h_1 & h_2 & h_3 \\ 0 & h_5 & 0 \\ \text{-}h_3 & h_8 & h_1 \end{bmatrix} \ , \tag{7}$$

where $h_1 = \cos\theta$, $h_2 = t_x$, $h_3 = \sin\theta$, $h_5 = 1 + t_y$, $h_8 = t_z$. Based on the homography constraint for points

$$[u_2, v_2, 1]_\times \mathtt{H_y} [u_1, v_1, 1]^\top = \mathtt{0} \ , \tag{8}$$

where $[u_2, v_2, 1]_\times$ is the skew symmetric matrix for the vector $[u_2, v_2, 1]^\top$, we have two linear constraints

$$h_1 v_2 - h_5 v_1 - h_3 u_1 v_2 + h_8 v_1 v_2 = 0 \ , \tag{9}$$

$$h_5 u_2 v_1 - h_1 u_1 v_2 - h_3 v_2 - h_2 v_1 v_2 = 0 \ . \tag{10}$$

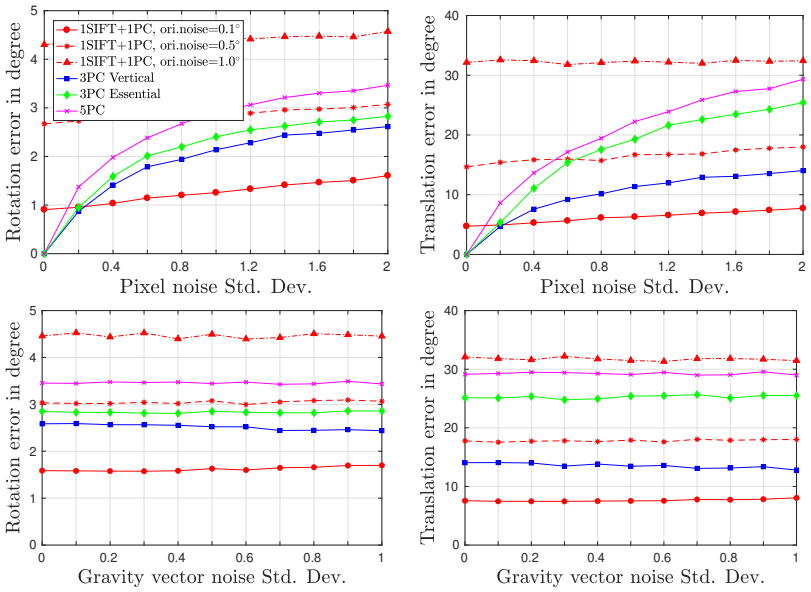

Fig. 2: Comparing the proposed 1SIFT+1PC solver with the three-point solver of Saurer et al. [5] when points are on a vertical plane (3PC Vertical), 3PC Essential [6] and 5PC [7]. The cameras undergo sideways motions. **Top**: Rotation and translation errors under increased pixel noise. **Bottom**: Rotation and translation errors under increased directional noise and constant image noise of 2 pixel standard deviation.

Using (6), the SIFT constraints are simplified to

$$h_8u_2s_1s_2 - h_3u_2s_2c_1 - h_8v_2s_1c_2 + h_3v_2c_1c_2 - h_2s_1s_2 - h_1s_2c_1 + h_5s_1c_2 = 0 \ , \quad (11)$$

$$h_7^2u_1^2q_2 + 2h_7h_8u_1v_1q_2 + h_8^2v_1^2q_2 + h_5h_7u_2q_1 - h_2h_7v_2q_1 + \quad (12)$$

$$h_1h_8v_2q_1 + 2h_7h_9u_1q_2 + 2h_8h_9v_1q_2 - h_1h_5q_1 + h_9^2q_2 = 0 \ ,$$

where $c_i = \cos(\varphi_i)$, $s_i = \sin(\varphi_i)$. Note that h_2 and h_8 can be formulated by $\{h_1, h_3, h_5\}$ using (9) and (10). Then substituting the formulation into (11), we obtain the following equation

$$\begin{aligned} &h_3v_2s_1s_2 + h_1v_2^2s_1c_2 - h_1v_1v_2c_1s_2 + h_1u_1v_2s_1s_2 - h_1u_2v_2s_1s_2 + \\ &h_3v_1v_2^2c_1c_2 - h_3u_1v_2^2s_1c_2 + h_3u_1u_2v_2s_1s_2 - h_3u_2v_1v_2c_1s_2 = 0 \ , \end{aligned} \quad (13)$$

where h_5 is also eliminated. Now equation (13) contains only two elements of $\mathtt{H}_y$, i.e., h_1 and h_3. Note that these two elements represent the rotation of the camera around the y-axis, i.e., $h_1 = \cos\theta$ and $h_3 = \sin\theta$. This means that $h_1^2 + h_3^2 = 1$ and, therefore, θ can be computed from (13). Since equation (13) does not contain the scale parameters of features q_1 and q_2, it follows that the camera rotation estimation is independent of the scale parameters of the features.

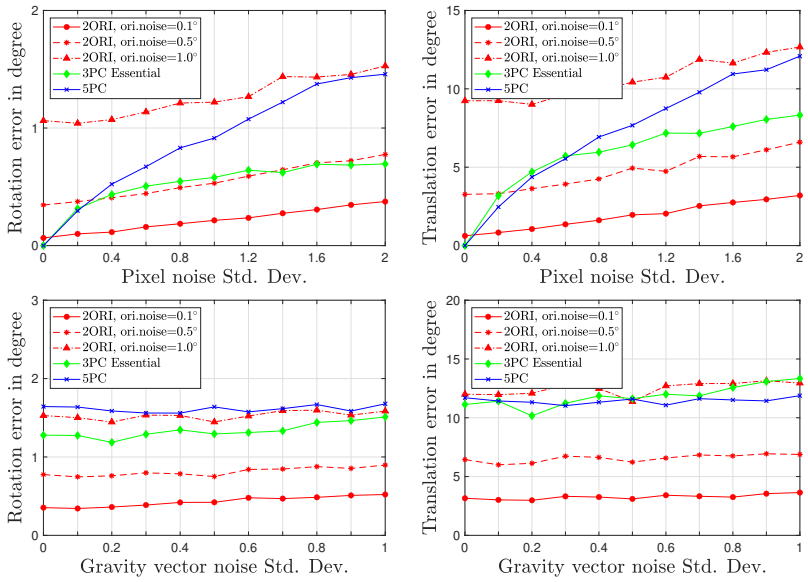

Fig. 3: Comparing the proposed 2ORI solver with the 3PC Essential [6] and 5PC [7]. The cameras undergo forward motions. **Top**: Rotation and translation errors under increased pixel noise. **Bottom**: Rotation and translation errors under increased directional noise and constant image noise of 2 pixel standard deviation.

References

1. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: 2011 International conference on computer vision, Ieee (2011) 2564–2571
2. Ding, Y., Yang, J., Ponce, J., Kong, H.: An efficient solution to the homography-based relative pose problem with a common reference direction. In: The IEEE International Conference on Computer Vision (ICCV). (2019)
3. Kukelova, Z., Bujnak, M., Pajdla, T.: Automatic generator of minimal problem solvers. In: European Conference on Computer Vision, Springer (2008) 302–315
4. Larsson, V., Åström, K., Oskarsson, M.: Efficient solvers for minimal problems by syzygy-based reduction. In: CVPR. Volume 2. (2017) 4
5. Saurer, O., Vasseur, P., Boutteau, R., Demonceaux, C., Pollefeys, M., Fraundorfer, F.: Homography based egomotion estimation with a common direction. IEEE transactions on pattern analysis and machine intelligence **39** (2017) 327–341
6. Fraundorfer, F., Tanskanen, P., Pollefeys, M.: A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In: European Conference on Computer Vision, Springer (2010) 269–282
7. Nistér, D.: An efficient solution to the five-point relative pose problem. IEEE transactions on pattern analysis and machine intelligence **26** (2004) 756–770